

09/1135403

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2003, 11:33:06 ; Search time 83 Seconds  
(without alignments)  
47.809 Million cell updates/sec

Title: SEQID2-NAT3  
Perfect score: 136  
Sequence: 1 RPNYSRRLLPKGVKHLKDFPILPGEI 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	93.4	56	23	ABJ04946
2	127	93.4	294	23	AAG79422
3	127	93.4	368	16	AAR73020
4	127	93.4	720	16	AAR74088
5	127	93.4	729	16	AAR74089
6	127	93.4	740	16	AAR76961
7	127	93.4	740	16	AAR76962
8	127	93.4	740	16	AAR73021
9	127	93.4	740	16	AAR74090

10	127	93.4	740	16	AAR76982	Human factor VIII
11	127	93.4	1383	18	AAW33227	Procoagulant-activ
12	127	93.4	1383	18	AAW33228	Procoagulant-activ
13	127	93.4	1383	18	AAW33229	Procoagulant-activ
14	127	93.4	1424	9	AAP80268	Modified factor VI
15	127	93.4	1424	10	AAP91169	Sequence of 740 Ar
16	127	93.4	1424	22	AAB48842	Mutant mature huma
17	127	93.4	1424	23	AAO18622	Human mature B-dom
18	127	93.4	1425	9	AAP80267	Modified factor VI
19	127	93.4	1438	21	AAB01262	B-domain deleted f
20	127	93.4	1440	12	AAR12971	Factor VIII:SQ. U
21	127	93.4	1445	23	ABG92540	LE B-Domain-delete
22	127	93.4	1447	23	ABG92541	5Arg B-domain-dele
23	127	93.4	1457	19	AAW46246	Human factor VIII
24	127	93.4	1457	19	AAW44372	Human Factor VIII
25	127	93.4	1457	20	AAW21675	Beta-domain delete
26	127	93.4	1459	22	AAE10827	Human factor VIII
27	127	93.4	1459	22	AAE10832	Human factor VIII
28	127	93.4	1459	22	AAE10833	Human factor VIII
29	127	93.4	1471	18	AAW23414	Human B-domain del
30	127	93.4	1471	22	AAB67959	Amino acid sequenc
31	127	93.4	1516	9	AAP80265	Modified factor VI
32	127	93.4	1661	18	AAW18670	Factor VIII-dB695-
33	127	93.4	2098	17	AAR86863	Factor-VIII. Homo
34	127	93.4	2332	8	AAP71726	Factor VIII:c varia
35	127	93.4	2332	8	AAP71727	Factor VIII:c varia
36	127	93.4	2332	8	AAP71728	Factor VIII:c varia
37	127	93.4	2332	8	AAP71729	Factor VIII:c varia
38	127	93.4	2332	14	AAR43257	Human Factor VIII.
39	127	93.4	2332	18	AAW33222	Procoagulant-activ
40	127	93.4	2332	18	AAW33223	Procoagulant-activ
41	127	93.4	2332	18	AAW33224	Procoagulant-activ
42	127	93.4	2332	18	AAW33225	Procoagulant-activ
43	127	93.4	2332	18	AAW33226	Procoagulant-activ
44	127	93.4	2332	19	AAW53483	Human factor VIII.
45	127	93.4	2332	19	AAW44132	Homo sapiens modif

ALIGNMENTS

RESULT 1

ABJ04946  
ID ABJ04946 standard; Protein; 56 AA.

XX  
AC ABJ04946;

XX  
DT 30-OCT-2002 (first entry)

XX  
DE A2 domain of fVII epitope SEQ ID NO 2.

XX  
KW Haemostatic; antibody inhibitor; factor VIII; T cell; immune response;

XX  
KW haemophilia A; acquired haemophilia; human factor VIII.

XX  
OS Homo sapiens.

XX  
PN WO200260917-A2.

XX  
PD 08-AUG-2002.

XX  
PF 30-NOV-2001; 2001WO-US44945.

XX  
PR 01-DEC-2000; 2000US-250430P.

XX  
PA (MINU ) UNIV MINNESOTA.

XX  
PI Conti-fine BM;

XX  
DR WPI; 2002-627462/67.

XX  
PT New peptides or their variants, useful for preventing, treating or

XX  
PT inhibiting aberrant or pathogenic production of antibodies specific for

XX  
PT factor VIII, particularly useful for treating hemophilia A or acquired



```

RESULT 4
AAR74088
ID AAR74088 standard; protein; 720 AA.
XX
AC AAR74088;
XX
DT 25-MAR-2003 (updated)
DT 04-NOV-1995 (first entry)
XX
DE Factor-VIII heavy chain N-terminal fragment.
XX
KW human: Factor VIII; heavy chain; N-terminal fragment;
KW thrombin cleavage; blood-clotting.
XX
OS Homo sapiens.
XX
PN WO9513300-A1.
XX
PD 18-MAY-1995.
XX
PF 10-NOV-1994; 94WO-DK00423.
XX
PR 12-NOV-1993; 93DK-0001280.
XX
PA (NOVO ) NOVO-NORDISK AS.
XX
PI Ezban Rasmussen M, Kjalke M;
DR WPI; 1995-194037/25.
XX
Factor VIII polypeptide(s) comprising a heavy chain shorter than native
Al-A2 domain - are easier to produce recombinantly and retain coagulant
activity, may be used to treat patients who have developed antibodies to
C-terminal epitope(s) of Factor VIII
XX
PS Claim 3; Page 24-26; 5lpp; English.
XX
The sequence represents N-terminal residues 1-720 of a human Factor-
VIII heavy chain. The sequence is shorter than the Al-A2 domain,
and is produced by treating a polypeptide containing the full Al-A2
domain of full-length Factor-VIII with a protease, e.g. thrombin.
The fragment has the same specific activity as full-length
Factor-VIII in a chromogenic assay, but activity is a factor of two
lower in a clotting assay, and the fragment is activated by thrombin
at a slower rate and to a lower level than fragments 1-740 (AAR74096).
1-729 (AAR74089) and plasma Factor-VIII. The fragment may be produced
recombinantly to reduce production costs and improve safety, and
production levels and stability are higher than for the full-length
form. The fragment may be used to treat patients who have developed
antibodies against epitopes in the C-terminal part of the heavy chain.
(Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 720 AA;
Query Match 93.4%; Score 127; DB 16; Length 720;
Best Local Similarity 96.0%; Pred. No. 5e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRRLPGVKVHLKDFPILPGEI 25
|| |||||
DB 484 RPLYSRRRLPGVKVHLKDFPILPGEI 508

RESULT 5
AAR74089
ID AAR74089 standard; protein; 729 AA.
XX
AC AAR74089;
XX
DT 25-MAR-2003 (updated)
DT 04-NOV-1995 (first entry)
XX
DE Factor-VIII heavy chain N-terminal fragment.

```

```

XX
KW human: Factor VIII; heavy chain; N-terminal fragment;
KW thrombin cleavage; blood-clotting.
XX
OS Homo sapiens.
XX
PN WO9513300-A1.
XX
PD 18-MAY-1995.
XX
PF 10-NOV-1994; 94WO-DK00423.
XX
PR 12-NOV-1993; 93DK-0001280.
XX
PA (NOVO ) NOVO-NORDISK AS.
XX
PI Ezban Rasmussen M, Kjalke M;
DR WPI; 1995-194037/25.
XX
Factor VIII polypeptide(s) comprising a heavy chain shorter than native
Al-A2 domain - are easier to produce recombinantly and retain coagulant
activity, may be used to treat patients who have developed antibodies to
C-terminal epitope(s) of Factor VIII
XX
PS Claim 2; Page 27-29; 5lpp; English.
XX
The sequence represents N-terminal residues 1-729 of a human Factor-
VIII heavy chain. The sequence is shorter than the Al-A2 domain,
and is produced by treating a polypeptide containing the full Al-A2
domain of full-length Factor-VIII with a protease, e.g. thrombin.
The fragment has the same coagulant specific activity as full-length
Factor-VIII in a chromogenic assay, and is activated by thrombin at
a similar rate. The fragment may be produced recombinantly to reduce
production costs and improve safety, and production levels and
stability are higher than for the full-length form. The fragment may
be used to treat patients who have developed antibodies against
epitopes in the C-terminal part of the heavy chain.
(Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 729 AA;
Query Match 93.4%; Score 127; DB 16; Length 729;
Best Local Similarity 96.0%; Pred. No. 5.1e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRRLPGVKVHLKDFPILPGEI 25
|| |||||
DB 484 RPLYSRRRLPGVKVHLKDFPILPGEI 508

RESULT 6
AAR76961
ID AAR76961 standard; protein; 740 AA.
XX
AC AAR76961;
XX
DT 09-MAR-1996 (first entry)
XX
DE Human Factor-VIII derivative.
XX
KW Factor-VIII; therapeutic; blood-clotting.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 692 /label= absent or Ala, Thr, Ser, Gly or Asp
FT Misc-difference 720 /label= absent or Gln, Ser, Thr, Val or Ala
FT Misc-difference 729 /label= absent or Val, Ala or Ile
XX

```



PN WO9518827-A1.  
XX 13-JUL-1995.  
PD  
XX 06-JAN-1995; 95WO-DK00008.  
PF  
XX 07-JAN-1994; 94DK-0000032.  
PR  
XX (NOVO ) NOVO-NORDISK AS.  
PA  
XX  
PI Ezban Rasmussen M, Nicolaisen EM, Persson E;  
XX WPI; 1995-255039/33.  
DR  
XX Novel factor VIII derivative used to treat haemophilia - and  
PT comprises a functional A2 domain containing a mutation at one or  
PT more Cys residues.  
PT  
XX Disclosure; Page 14-17; 30pp; English.  
PS  
XX The new Factor-VIII derivative comprises a functional A2 domain in  
CC which Cys-692 is deleted or replaced with another amino acid  
CC residue, preferably Ser (see AAR76962). Alternatively, Glu-720 and/or  
CC Tyr-729 are deleted or substituted with various amino acids (as in  
CC the Features). The new derivative has the same activity as the wild-  
CC type Factor-VIII but with improved stability (the activity is  
CC maintained for a longer period compared to the rapid decline of the  
CC activity of wt Factor-VIII). The new derivative can be used in a  
CC composition for treating diseases caused by an absence or deficiency  
CC of Factor-VIII, especially haemophilia.  
XX  
SQ Sequence 740 AA;  
  
Query Match 93.4%; Score 127; DB 16; Length 740;  
Best Local Similarity 96.0%; Pred. No. 5.2e-10;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25  
II |  
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508  
  
RESULT 7  
AAR76962  
ID AAR76962 standard; protein; 740 AA.  
XX  
AC AAR76962;  
XX  
DT 09-MAR-1996 (first entry)  
XX  
DE Human Factor-VIII derivative.  
XX  
KW Factor-VIII; therapeutic; blood-clotting.  
XX  
OS Homo sapiens.  
PH Key Location/Qualifiers  
FT Misc-difference 692 /label= Cys substituted by Ser  
FT Misc-difference 720 /label= absent or Gln, Ser, Thr, Val or Ala  
FT Misc-difference 729 /label= absent or Val, Ala or Ile  
FT  
XX WO9518827-A1.  
PN  
XX 13-JUL-1995.  
PD  
XX 06-JAN-1995; 95WO-DK00008.  
PF  
XX 07-JAN-1994; 94DK-0000032.  
PR  
XX (NOVO ) NOVO-NORDISK AS.  
PA

XX Ezban Rasmussen M, Nicolaisen EM, Persson E;  
PI WPI; 1995-255039/33.  
XX  
DR Novel factor VIII derivative used to treat haemophilia - and  
XX comprises a functional A2 domain containing a mutation at one or  
PT more Cys residues.  
PT  
XX Disclosure; Page 18-20; 30pp; English.  
PS  
XX The new Factor-VIII derivative comprises a functional A2 domain in  
CC which Cys-692 is replaced with Ser. For other (less preferred)  
CC substitutions at this site, see AAR76961. Alternatively, Glu-720 and/or  
CC Tyr-729 are deleted or substituted with various amino acids (as in  
CC the Features). The new derivative has the same activity as the wild-  
CC type Factor-VIII but with improved stability (the activity is  
CC maintained for a longer period compared to the rapid decline of the  
CC activity of wt Factor-VIII). The new derivative can be used in a  
CC composition for treating diseases caused by an absence or deficiency  
CC of Factor-VIII, especially haemophilia.  
XX  
SQ Sequence 740 AA;  
  
Query Match 93.4%; Score 127; DB 16; Length 740;  
Best Local Similarity 96.0%; Pred. No. 5.2e-10;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25  
II |  
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508  
  
RESULT 8  
AAR73021  
ID AAR73021 standard; peptide; 740 AA.  
XX  
AC AAR73021;  
XX  
DT 25-MAR-2003 (updated)  
DI 21-NOV-1995 (first entry)  
XX  
DE Human Factor-VIII N-terminal fragment.  
XX  
KW Factor-VIII; blood-clotting; blood; coagulant; haemophilia-A.  
XX  
OS Homo sapiens.  
XX  
PN WO9513301-A1.  
XX  
PD 18-MAY-1995.  
XX  
PF 10-NOV-1994; 94WO-DK00424.  
XX  
PR 12-NOV-1993; 93DK-0001281.  
XX  
PA (NOVO ) NOVO-NORDISK AS.  
XX  
PI Persson E;  
XX  
XX WPI; 1995-194038/25.  
DR  
XX Crosslinked Factor VIII polypeptide which is stable - is prepd. using  
PT bis(sulphosuccinimydyl) suberate or disuccinimydyl suberate in the  
PT presence of polysorbate 80 to produce a coagulant with long lasting  
PT activity  
XX  
PS Disclosure; Page 21; 36pp; English.  
XX  
CC This is the N-terminal fragment of human Factor-VIII which may be  
CC crosslinked resulting in increased stability and retention of high  
CC activity over extended periods of time after activation by thrombin.  
CC The polypeptide is used to prevent or treat diseases caused by the

CC absence or deficiency of Factor-VIII in a subject such as  
CC haemophilia.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
XX

SQ Sequence 740 AA;

Query Match 93.4%; Score 127; DB 16; Length 740;  
Best Local Similarity 96.0%; Pred. No. 5.2e-10;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLLPKGVKHLKDFPILPGEI 25  
|| |||||  
Db 484 RPLYSRRLLPKGVKHLKDFPILPGEI 508

RESULT 9  
AAR74090  
ID AAR74090 standard; protein; 740 AA.  
XX  
AC AAR74090;  
XX  
DT 25-MAR-2003 (updated)  
DT 04-NOV-1995 (first entry)  
XX  
DE Factor-VIII heavy chain N-terminal fragment.  
XX  
KW human; Factor VIII; heavy chain; N-terminal fragment;  
KW thrombin cleavage; blood-clotting.  
XX  
OS Homo sapiens.  
XX  
PN WO9513300-A1.  
XX  
PD 18-MAY-1995.  
XX  
PF 10-NOV-1994; 94WO-DK00423.  
XX  
PR 12-NOV-1993; 93DK-0001280.  
XX  
PA (NOVO ) NOVO-NORDISK AS.  
XX  
PI Ezban Rasmussen M, Kjalke M;  
XX  
DR WPI; 1995-194037/25.  
XX  
PT Factor VIII polypeptide(s) comprising a heavy chain shorter than native  
PT A1-A2 domain - are easier to produce recombinantly and retain coagulant  
PT activity, may be used to treat patients who have developed antibodies to  
PT C-terminal epitope(s) of Factor VIII  
XX  
PS Disclosure; Page 30-32; 51pp; English.  
XX

CC The sequence represents N-terminal residues 1-740 of a human Factor-  
CC VIII heavy chain. The sequence contains entire A1 and A2 domains,  
CC and truncated forms (1-720 in (AAR74088) and 1-729 in (AAR74090))  
CC may be produced by treatment with a protease, e.g. thrombin. The  
CC C-terminally truncated fragments have the same coagulant specific  
CC activity as full-length Factor-VIII, and may be produced  
CC recombinantly to reduce production costs and improve safety, giving  
CC higher production levels and stability than for the full-length form.  
CC The fragments may be used to treat patients who have developed  
CC antibodies against epitopes in the C-terminal part of the heavy chain.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 740 AA;

Query Match 93.4%; Score 127; DB 16; Length 740;  
Best Local Similarity 96.0%; Pred. No. 5.2e-10;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLLPKGVKHLKDFPILPGEI 25  
|| |||||  
Db 484 RPLYSRRLLPKGVKHLKDFPILPGEI 508

RESULT 10  
AAR76982  
ID AAR76982 standard; protein; 740 AA.  
XX  
AC AAR76982;  
XX  
DT 21-FEB-1996 (first entry)  
XX  
DE Human factor VIII A2-domain derivative.  
XX  
KW Factor VIII; human; haemophilia; thrombin; protein C; plasmin;  
KW serine protease; recombination; therapy; deficiency.  
XX  
OS Homo sapiens.  
XX

FH Key Location/Qualifiers  
FT Misc-difference 720  
FT /label= Gln, Ser, Thr, Val, Ala  
FT Misc-difference 729  
FT /label= Val, Ala, Ile  
XX  
PN WO9518828-A1.  
XX  
PD 13-JUL-1995.  
XX  
PF 06-JAN-1995; 95WO-DK000009.  
XX  
PR 07-JAN-1994; 94DK-0000031.  
XX  
PA (NOVO ) NOVO-NORDISK AS.  
XX  
PI Ezban Rasmussen M, Nicolaisen EM, Persson E;  
XX  
DR WPI; 1995-255040/33.  
XX

PT Novel factor VIII derivative with resistance to enzymatic cleavage -  
PT and comprises a functional A2 domain where Glu720 and/or Tyr729 is  
PT deleted/substituted.  
XX  
PS Claim 1; Page 11-14; 26pp; English.  
XX

CC This sequence represents the A2 domain of a human factor VIII derivative.  
CC Factor VIII is a large glycoprotein which is present in plasma at low  
CC concentrations. Factor VIII is an essential part of the clotting  
CC reaction in response to a wound. Factor VIII is susceptible to cleavage  
CC by thrombin, activated protein C, plasmin, and other serine proteases.  
CC Full length factor VIII consists of three repeats of the A-domain, a  
CC B-domain and 2 repeats of the C-domain. Active factor VIII has the A1  
CC domain cleaved off. Factor VIII is too unstable for use in recombinant  
CC techniques. Factor VIII containing this sequence has improved stability  
CC and shows resistance against enzymatic activity present in mammalian  
CC cells. This means that factor VIII containing this sequence can be used  
CC in recombinant techniques. A factor VIII derivative can be used for  
CC treating diseases caused by an absence or deficiency of factor VIII (in  
CC the same way as normal factor VIII) e.g. haemophilia. The advantage with  
CC using a recombinant factor VIII also includes no need for lots of donors  
CC in order to get a sufficient amount. Also, there is no long purification  
CC process, and there is no risk of transmission of blood-borne diseases  
CC such as HIV.  
XX  
SQ Sequence 740 AA;

Query Match 93.4%; Score 127; DB 16; Length 740;  
Best Local Similarity 96.0%; Pred. No. 5.2e-10;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLLPKGVKHLKDFPILPGEI 25  
|| |||||  
Db 484 RPLYSRRLLPKGVKHLKDFPILPGEI 508





FT                    380..554                    between domains A2 and A3"  
FT                    /note= "plastocyanin-like domain 3"  
FT                    564..711  
FT                    /note= "plastocyanin-like domain 4"  
FT                    746..1073  
FT                    /note= "A3 domain"  
FT                    1073..1221  
FT                    /note= "C1 domain"  
FT                    1226..1378  
FT                    /note= "C2 domain"  
FT                    372..373  
FT                    /note= "by thrombin"  
FT                    153..179  
FT                    /note= "probable"  
FT                    528..554  
FT                    /note= "probable"  
FT                    336  
FT                    /label= R336I  
FT                    /note= "wild type Arg replaced with Ile"  
FT                    562  
FT                    /label= R562K  
FT                    /note= "wild type Arg replaced with Lys"  
FT                    740  
FT                    /label= R740A  
FT                    /note= "wild type Arg replaced with Ala"  
XX  
PN                    WO9740145-A1.  
XX  
PD                    30-OCT-1997.  
XX  
PF                    24-APR-1997;    97WO-US06563.  
XX  
PR                    15-MAY-1996;    96US-0017785.  
PR                    24-APR-1996;    96US-0016117.  
XX  
PA                    (UNMI ) UNIV MICHIGAN.  
XX  
PI                    Ariano K,   Kaufman RJ,   Pipe SW;  
XX  
XX                    WPI; 1997-535830/49.  
DR  
PT                    Modified human pro-coagulant active factor VIII - can be  
PT                    administered to haemophiliacs, i.e. factor VIII replacement therapy  
XX  
PS                    Claim i8; Page -; 57pp; English.  
XX  
CC                    The present sequence represents a novel pro-coagulant active factor  
CC                    VIII (FVIII) mutant protein, comprising a deletion of the B domain and  
CC                    von Willebrand factor binding site, mutations R336I, R562K and R740A and  
CC                    an addition of an amino acid sequence spacer between the A2 and A3  
CC                    domains. Factor VII, along with calcium and phospholipid, acts as a  
CC                    cofactor for factor IXA, when it converts factor X to the activated form  
CC                    (factor XA). FVIII is the coagulation factor deficient in the  
CC                    X-chromosome-linked bleeding disorder haemophilia A. Several other  
CC                    mutant FVIII proteins have also been created (see AAW33222-29). The FVIII  
CC                    mutant F309S (AAW33225) is capable of recombinant secretion at higher  
CC                    levels than typically obtained with wild type FVIII and retains  
CC                    pro-coagulant activity. The FVIII mutant R336I (AAW33222) and R562K  
CC                    (AAW33223) are resistant to activated protein C (APC) cleavage. The FVIII  
CC                    mutant comprising a deletion of the B domain and von Willebrand factor  
CC                    binding site, a mutation at Arg740 and an addition of an amino acid  
CC                    sequence spacer between the A2 and A3 domains can form a more stable  
CC                    configuration, and have an approximate 5-fold increase in specific  
CC                    activity compared to purified wild type FVIII, while increasing their  
CC                    binding affinity to von Willebrand factor improves their stability.  
CC                    The FVIII proteins can be administered to haemophiliacs, i.e. FVIII  
CC                    replacement therapy, while the nucleic acid molecule can be used for  
CC                    gene therapy.  
CC                    note: this sequence does not appear in the specification; it was created  
CC                    using sequences from the given references.  
XX  
SQ                    Sequence    1383 AA;

Query Match                    93.4%;    Score 127;    DB 18;    Length 1383;  
Best Local Similarity    96.0%;    Pred. No. 1e-09;  
Matches    24;    Conservative    0;    Mismatches    1;    Indels    0;    Gaps    0;  
  
QY                    1 RPNYSRRRLPKGVKHLKDFPILPGEI 25  
                  || |||||  
Db                    484 RPLYSRRLPKGVKHLKDFPILPGEI 508  
  
RESULT 13  
AAW33229  
ID    AAW33229 standard; protein; 1383 AA.  
XX  
AC    AAW33229;  
XX  
DT    30-APR-1998    (first entry)  
XX  
DE    Procoagulant-active human factor VIII:C (FVIII) mutant protein.  
XX  
KW    Pro-coagulant active factor VIII; FVIII; haemophilia A;  
KW    recombinant secretion; pro-coagulant activity; resistance;  
KW    activated protein C cleavage; APC; B domain; A2 domain; A3 domain;  
KW    von Willebrand factor binding site; binding affinity;  
KW    FVIII replacement therapy.  
XX  
OS    Synthetic.  
OS    Homo sapiens.  
XX  
FH    Key                    Location/Qualifiers  
FT    Region                1..346  
FT                    /note= "factor VIIIA heavy chain"  
FT    Region                741..1383  
FT                    /note= "factor VIIIA light chain"  
FT    Domain                1..329  
FT                    /note= "A1 domain"  
FT    Domain                1..179  
FT                    /note= "plastocyanin-like domain 1"  
FT    Domain                187..329  
FT                    /note= "plastocyanin-like domain 2"  
FT    Domain                380..711  
FT                    /note= "A2 domain"  
FT    Misc\_feature           711..746  
FT                    /note= "a spacer of the sequence  
FT                    SFQNSRHPSTRQKQFNATTPENDIEKTDPMF  
FT                    AHRTMPKIQNVSSDLMLL is inserted  
FT                    between domains A2 and A3"  
FT  
FT    Domain                380..554  
FT                    /note= "plastocyanin-like domain 3"  
FT    Domain                564..711  
FT                    /note= "plastocyanin-like domain 4"  
FT    Domain                746..1073  
FT                    /note= "A3 domain"  
FT    Domain                1073..1221  
FT                    /note= "C1 domain"  
FT    Domain                1226..1378  
FT                    /note= "C2 domain"  
FT    Cleavage-site           372..373  
FT                    /note= "by thrombin"  
FT    Disulfide-bond          153..179  
FT                    /note= "probable"  
FT    Disulfide-bond          528..554  
FT                    /note= "probable"  
FT    Misc-difference          309  
FT                    /label= F309S  
FT                    /note= "wild type Phe replaced with Ser"  
FT    Misc-difference          740  
FT                    /label= R740A  
FT                    /note= "wild type Arg replaced with Ala"  
XX  
XX                    WO9740145-A1.  
PN  
PD                    30-OCT-1997.





PT Factor VIII:C and a promoter  
XX  
PS Disclosure; Fig 1; 32pp; English.  
XX  
CC Arg-740 of the carboxyl terminus of the H chain is  
CC directly bonded by a peptide bond to Glu-1649 of the amino terminus of  
CC L chain. A pref. expression vector used to transform animal cell so  
CC that they produce human Factor VIII:Cis plasmid Ad.RE.neo.  
CC The expression vector has at least one promoter upstream of AAN90654.  
CC The transformants can constantly and continuously produce human Factor  
CC VIII:C in high yield on a commercial scale. The human Factor VIII:C so  
CC produced is considered to corresp. to the smallest species of active and  
CC intact Factor VIII:C molecules in the human blood plasma. It is useful  
CC for treating haemophilia A patients.  
CC (Updated on 25-MAR-2003 to correct PR field.)  
CC (Updated on 25-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 1424 AA;  
  
Query Match 93.4%; Score 127; DB 10; Length 1424;  
Best Local Similarity 96.0%; Pred. No. 1e-09;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 RPNYSRRRLPKGKHLKDFPILPGEI 25  
|| |||||  
Db 484 RPLYSRRRLPKGKHLKDFPILPGEI 508  
  
Search completed: October 17, 2003, 11:35:08  
Job time : 83 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run on:           October 17, 2003, 11:33:11 ; Search time 40 Seconds  
                  (without alignments)  
                  60.105 Million cell updates/sec

Title:            SEQID2-NAT3  
Perfect score:   136  
Sequence:        1 RPNYSRRLPKGVKHLKDFPILPGEI 25

Scoring table:   BLOSUM62  
                  Gapop 10.0 , Gapext 0.5

Searched:        283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:       283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
                  Maximum Match 100%  
                  Listing first 45 summaries

Database :       PIR\_76: \*  
                  1: pirl: \*  
                  2: pirl: \*  
                  3: pirl: \*  
                  4: pirl: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	127	93.4	2351	1 EZHU	coagulation factor
2	94	69.1	2319	2 A47004	coagulation factor
3	75	55.1	2133	2 T42763	coagulation factor
4	51	37.5	1607	2 T04583	TMV resistance pro
5	50	36.8	958	2 A82583	conserved hypothet
6	49	36.0	412	2 G71070	hypothetical prote
7	48.5	35.7	400	2 JG4591	alpha-1.3 fucosylt
8	48.5	35.7	405	2 B36340	alpha(1.3)-fucosyl
9	48.5	35.7	433	2 A57596	alpha-1.3-fucosylt
10	48.5	35.7	1253	2 T45787	disease resistance
11	48	35.3	156	2 B83164	conserved hypothet
12	48	35.3	264	2 AC2515	hypothetical prote
13	48	35.3	600	2 B45642	DNA-directed DNA p
14	48	35.3	1690	2 T40847	probable rRNA biog
15	47	34.6	444	2 S54011	1-aminocyclopropan
16	46	33.8	308	2 S67657	hypothetical prote
17	46	33.8	348	2 G71681	hypothetical prote
18	46	33.8	467	2 T38353	serine hydroxymeth
19	46	33.8	480	2 A99164	hypothetical prote
20	46	33.8	519	2 S77572	oligopeptide trans
21	46	33.8	885	2 T38387	hypothetical prote
22	46	33.8	1001	2 C88779	protein T20D3.9 [i
23	46	33.8	1023	2 A59431	KIAA0013 protein [
24	46	33.8	1038	2 T25033	hypothetical prote
25	45.5	33.5	161	2 F75633	hypothetical prote
26	45.5	33.5	246	2 H82553	3-demethylubiquino
27	45.5	33.5	554	2 B90679	3-(3-hydroxyphenyl
28	45.5	33.5	554	2 F85529	3-(3-hydroxyphenyl
29	45.5	33.5	554	2 C64762	probable monooxyge

30	45	33.1	61	2 AC2890	hypothetical prote
31	45	33.1	204	2 AB2365	hypothetical prote
32	45	33.1	438	2 S76347	UDP-N-acetylglucos
33	45	33.1	444	2 T15310	hypothetical prote
34	45	33.1	1690	1 CGHU1B	collagen alpha 4(I
35	45	33.1	1691	1 CGHU6B	collagen alpha 5(I
36	45	33.1	1784	2 T10532	gag-pol polyprotei
37	45	33.1	2054	2 T07584	hypothetical prote
38	44.5	32.7	331	2 F89771	lipoprotein [impor
39	44.5	32.7	485	2 G81308	IMP dehydrogenase
40	44.5	32.7	788	2 A72330	phenylalanine-tRNA
41	44.5	32.7	906	2 JC5963	stable tubule only
42	44	32.4	68	2 G69102	hypothetical prote
43	44	32.4	88	2 T17715	hypothetical prote
44	44	32.4	130	1 R3LV11	ribosomal protein
45	44	32.4	161	2 S33491	hypothetical prote

ALIGNMENTS

RESULT 1

EZHU  
coagulation factor VIII precursor [validated] - human  
N:Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant c  
C:Species: Homo sapiens (man)  
C>Date: 28-Aug-1985 \*sequence\_revision 28-Aug-1985 \*text\_change 08-Dec-2000  
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445;  
R:Gitschier, J.; Wood, W.I.  
Hum. Mol. Genet. 1, 199-200, 1992  
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.  
A:Reference number: I54318; MUID:93265012; PMID:1303178  
A:Accession: I54318  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1921, 'S', 1923-2351 <RES>  
A:Cross-references: GB:M88648; NID:g182381; PIDN:AAA52420.1; PID:g182383  
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; See  
Nature 312, 330-337, 1984  
A:Title: Expression of active human factor VIII from recombinant DNA clones.  
A:Reference number: A00525; MUID:85061548; PMID:6436526  
A:Accession: A00525  
A:Molecule type: mRNA  
A:Residues: 1-2351 <WOO>  
A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179  
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.  
S., D.N.; Hewick, R.M.  
Nature 312, 342-347, 1984  
A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.  
A:Reference number: I58059; MUID:85061550; PMID:6438528  
A:Accession: I58059  
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-74, 'V', 76-1259, 'E', 1261-2351 <RE2>  
A:Cross-references: GB:K01740; NID:g182802; PIDN:AAA52484.1; PID:g182803  
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.;  
.B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.  
DNA 4, 333-349, 1985  
A:Title: Characterization of the polypeptide composition of human factor VIII:C and  
A:Reference number: A23584; MUID:86081164; PMID:3935400  
A:Accession: A23584  
A:Molecule type: mRNA  
A:Residues: 1-2351 <TRU>  
A:Cross-references: GB:M14113; NID:g182817; PIDN:AAA52485.1; PID:g182818  
R:Eaton, D.; Rodriguez, H.; Vehar, G.A.  
Biochemistry 25, 505-512, 1986  
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleava-  
ity.  
A:Reference number: A26174; MUID:86159740; PMID:3082357  
A:Accession: A26174  
A:Molecule type: protein  
A:Residues: 20-36;392-399, 'X', 401-402;1668-1678;1709-1722, 'D', 1723-1725;1741-1755 <E  
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.





Query Match 55.1%; Score 75; DB 2; Length 2133;  
Best Local Similarity 83.3%; Pred. No. 0.014;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 RLPKGVKHLKDFPILPGE 24  
||| ||||| |||||  
DB 509 RLLKGWKHLKDWPILPGE 526

RESULT 4  
T04583  
TMV resistance protein N homolog F23E13.30 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 17-Nov-2000  
C;Accession: T04583; T05507  
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.  
submitted to the Protein Sequence Database, March 1998  
A;Reference number: Z15378  
A;Accession: T04583  
A;Molecule type: DNA  
A;Residues: 1-1607 <BEV>  
A;Cross-references: EMBL:AL022141  
A;Experimental source: cultivar Columbia; BAC clone F23E13  
R;Bevan, M.; Wedler, H.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X.; Schue-  
submitted to the Protein Sequence Database, April 1998  
A;Reference number: Z15418  
A;Accession: T05507  
A;Molecule type: DNA  
A;Residues: 1448-1607 <BE2>  
A;Cross-references: EMBL:AL022373  
A;Experimental source: cultivar Columbia; BAC clone T19K4  
C;Genetics:  
A;Map position: 4  
A;Introns: 193/2; 238/2; 556/2; 930/3; 1029/3; 1287/3  
A;Note: F23E13.30; T19K4.270

Query Match 37.5%; Score 51; DB 2; Length 1607;  
Best Local Similarity 66.7%; Pred. No. 42;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 NYSRRLPKGVKHLKD i7  
||| ||||| |||  
DB 973 HYSRLPKGLKFLPD 987

RESULT 5  
A82583  
conserved hypothetical protein XF2237 [imported] - Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C;Accession: A82583  
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A;Reference number: A82515; MUID:20365717; PMID:10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: A82583  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-958 <SIM>  
A;Cross-references: GB:AE004036; GB:AE003849; NID:g9107384; PIDN:AAF85036.1; GSPDB:GN001  
A;Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A-  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H-  
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm-  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig-  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E-  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A-  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv-  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF2237

Query Match 36.8%; Score 50; DB 2; Length 958;  
Best Local Similarity 47.1%; Pred. No. 34;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 RRLPKGVKHLKDFPILP 22  
||:||||: | |:  
DB 124 RRIPKGIQETPDITIIP 140

RESULT 6  
G71070  
hypothetical protein PH1259 - Pyrococcus horikoshii  
C;Species: Pyrococcus horikoshii  
C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
C;Accession: G71070  
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se-  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogi-  
DNA Res. 5, 55-76, 1998  
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophil;  
A;Reference number: A71000; MUID:98344137; PMID:9679194  
A;Accession: G71070  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-412 <KAW>  
A;Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30361.1; PID:g3257678  
A;Experimental source: strain OT3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa-  
C;Genetics:  
A;Gene: PH1259  
C;Superfamily: hypothetical protein HI0333

Query Match 36.0%; Score 49; DB 2; Length 412;  
Best Local Similarity 58.8%; Pred. No. 19;  
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 PNYSRRLPKGVKHLKDF 18  
| | | | | | | | | |  
DB 140 PIFSERTPKYLKALKDF 156

RESULT 7  
JC4591  
alpha-1,3 fucosyltransferase (EC 2.4.1.-) - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 16-Apr-1996 #sequence\_revision 24-May-1996 #text\_change 24-Nov-1999  
C;Accession: JC4591  
R;Ozawa, M.; Muramatsu, T.  
J. Biochem. 119, 302-308, 1996  
A;Title: Molecular cloning and expression of a mouse alpha-1,3 fucosyltransferase ge-  
A;Reference number: JC4591; MUID:97037075; PMID:8882722  
A;Accession: JC4591  
A;Molecule type: mRNA  
A;Residues: 1-400 <OZA>  
A;Cross-references: DDBJ:D63379  
A;Experimental source: Embryonal carcinoma F9 cells  
C;Superfamily: galactoside 3(4)-L-fucosyltransferase  
C;Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane pro-  
F;1-23/Domain: intracellular #status predicted <INT>  
F;24-49/Domain: transmembrane #status predicted <TRM>  
F;64,185/Binding site: carbohydrate (Asn) (covalent) #status predicted  
Query Match 35.7%; Score 48.5; DB 2; Length 400;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 RPNYSRRLPKGV-KHLKDFP 19

Db 314 RANYERFVPRGAFIHVDFFP 333

RESULT 8

B36340

alpha(1,3)-fucosyltransferase (EC 2.4.1.1-) 4 precursor [validated] - human

N;Alternate names: CD15; ELAM-1 ligand fucosyltransferase (ELFT); FCT3A; FUC-TIV; myeloid

C;Species: Homo sapiens (man)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C;Accession: B36340; A36340; A40976; A41202

R;Goelz, S.E.; Hession, C.; Goff, D.; Griffiths, B.; Tizard, R.; Newman, B.; Chi-Rosso, C.

Cell 63, 1349-1356, 1990

A;Title: ELFT: a gene that directs the expression of an ELAM-1 ligand.

A;Reference number: A36340; MUID:91084863; PMID:1702034

A;Accession: B36340

A;Molecule type: mRNA

A;Residues: 1-405 <GOE1>

A;Cross-references: GB:M58596; NID:gl82068; PIDN:AAA63172.1; PID:gl82069

A;Accession: A36340

A;Molecule type: mRNA

A;Residues: 'MRRLWGAAKPKSGAGWEKEWAPQEAAPGAWSGRLGPR', 'SGRKGRAVPGWASWPAHLAALAAARHLGGAGQ

A;Cross-references: GB:M58597; NID:gl82070; PIDN:AAA63173.1; PID:gl82071

A;Note: the codon used as an initiator for this translation is not in a good context for

R;Low, J.B.; Kukowska-Latallo, J.F.; Nair, R.P.; Larsen, R.D.; Marks, R.M.; Machter, B.A.

J. Biol. Chem. 266, 17467-17477, 1991

A;Title: Molecular cloning of a human fucosyltransferase gene that determines expression

A;Reference number: A40976; MUID:91373370; PMID:1716630

A;Accession: A40976

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-86, 'P', 88-405 <LOW>

A;Cross-references: GB:M65030; NID:gl82791; PIDN:AAA92977.1; PID:gl236720

R;Kumar, R.; Potvin, B.; Muller, W.A.; Stanley, P.

J. Biol. Chem. 266, 21777-21783, 1991

A;Title: Cloning of a human alpha(1,3)-fucosyltransferase gene that encodes ELFT but does

A;Reference number: A41202; MUID:92042084; PMID:1718983

A;Accession: A41202

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-240, 'D', 242-400 <KUM>

A;Cross-references: GB:S65161; NID:g239005; PIDN:AAB20349.1; PID:g239006

C;Genetics:

A;Gene: GDB:FUT4; CD15; FCT3A; FUC-TIV

A;Cross-references: GDB:131563; OMIM:104230

A;Map position: 11q21-11q21

C;Superfamily: galactoside 3(4)-L-fucosyltransferase

C;Keywords: glycoprotein; glycosyltransferase; hexosyltransferase

F;1-48/Domain: signal sequence #status predicted <SIG>

F;49-405/Product: alpha(1,3)-fucosyltransferase 4 #status predicted <MAT>

F;91,190/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 35.7%; Score 48.5; DB 2; Length 405;

Best Local Similarity 50.0%; Pred. No. 23;

Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 1 RPNYSRRLPKGV-KHLKDFP 19

Db 319 RANYERFVPRGAFIHVDFFP 338

RESULT 9

A57596

alpha-1,3-fucosyltransferase Fuc-TIV (EC 2.4.1.1-) - mouse

N;Alternate names: ELAM-1 ligand fucosyltransferase homolog

C;Species: Mus musculus (house mouse)

C;Date: 08-Feb-1996 #sequence\_revision 08-Feb-1996 #text\_change 11-Jan-2000

C;Accession: A57596

R;Gersten, K.M.; Natsuka, S.; Trinchera, M.; Petryniak, B.; Kelly, R.J.; Hiraiwa, N.; Je

J. Biol. Chem. 270, 25047-25056, 1995

A;Title: Molecular cloning, expression, chromosomal assignment, and tissue-specific exp

erase.

A;Reference number: A57596; MUID:96027607; PMID:7559635

A;Accession: A57596

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-433 <GER>

A;Cross-references: GB:U33457; NID:gl039426; PIDN:AAC52269.1; PID:gl039427

C;Superfamily: galactoside 3(4)-L-fucosyltransferase

C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 35.7%; Score 48.5; DB 2; Length 433;

Best Local Similarity 50.0%; Pred. No. 24;

Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 1 RPNYSRRLPKGV-KHLKDFP 19

Db 347 RANYERFVPRGAFIHVDFFP 366

RESULT 10

T45787

disease resistance-like protein - Arabidopsis thaliana

N;Alternate names: protein F26013.200

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000

C;Accession: T45787

R;Delseny, M.; Berger, C.; Cooke, R.; Greillett, F.; Laudie, M.; Mewes, H.W.; Lemcke, I

submitted to the Protein Sequence Database, December 1999

A;Reference number: 223013

A;Accession: T45787

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1253 <DEL>

A;Cross-references: EMBL:AL133452

A;Experimental source: cultivar Columbia; BAC clone F26013

C;Genetics:

A;Map position: 3

A;Introns: 138/2; 490/3; 589/3; 1047/3

A;Note: F26013.200

Query Match 35.7%; Score 48.5; DB 2; Length 1253;

Best Local Similarity 39.3%; Pred. No. 76;

Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

Qy 3 NYSRRLPKGVK-----HLKDFPIL 21

Db 533 HYALHLPKGVKSLPEELRLHWEHFPLL 560

RESULT 11

B83164

conserved hypothetical protein PA3847 [imported] - Pseudomonas aeruginosa (strain PA

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: B83164

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; I

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: B83164

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-156 <STO>

A;Cross-references: GB:AE004802; GB:AE004091; NID:g9950021; PIDN:AAG07234.1; GSPDB:G

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA3847

Query Match 35.3%; Score 48; DB 2; Length 156;

Best Local Similarity 36.4%; Pred. No. 9.7;

Matches 8; Conservative 6; Mismatches 8; Indels 6; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILP 22

```
DB          33  R2LLSMSVPPGGIRSLDCEVVP 54
|| | | | | | | | | | | | | | | | |
RESULT 12
AC2515
hypothetical protein alr7299 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120al
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2515
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2515
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-264 <KUR>
A:Cross-references: GB:BA000020; PIDN:BAB78383.1; PID:gl7135837; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr7299
A:Genome: plasmid

Query Match          35.3%; Score 48; DB 2; Length 264;
Best Local Similarity 58.8%; Pred. No. 17;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY          8  LPKGVKHLKDFPILPGE 24
| | | | | | | | | |
DB          232  LTAGLKHLKSFDFYLPVE 248

RESULT 13
B46642
DNA-directed DNA polymerase (EC 2.7.7.7) alpha/DNA primase (EC 2.7.7.-) complex 68k chain
C:Species: Mus musculus (house mouse)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
C:Accession: B46642
R:Miyaazawa, H.; Izumi, M.; Tada, S.; Takada, R.; Masutani, M.; Ui, M.; Hanaoka, F.
J. Biol. Chem. 268, 8111-8122, 1993
A:Title: Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase al
A:Reference number: A46642; MUID:93216788; PMID:8463324
A:Accession: B46642
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-600 <MIY>
A:Cross-references: GB:D13546; NID:g303658; PIDN:BAA02746.1; PID:g303659
A:Experimental source: FM3A cells
A:Note: sequence extracted from NCBI backbone (NCBIN:129148, NCBI:P:129149)
C:Keywords: nucleotidyltransferase

Query Match          35.3%; Score 48; DB 2; Length 600;
Best Local Similarity 26.1%; Pred. No. 41;
Matches 6; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY          3  NYSRRLPKGVKHLKDFPILPGEI 25
| | | | | | | | | |
DB          279  SYGAQIPVDLSLKEYSLFPGQV 301

RESULT 14
T40847
probable rRNA biogenesis protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T40847
R:Hilbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barreil, B.G.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21951
A:Accession: T40847
```

```
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1690 <HIL>
A:Cross-references: EMBL:AL031740; PIDN:CAA21087.1; GSPDB:GN00068; SPDB:SPCC1183.07
A:Experimental source: strain 972h-; cosmid cll183
C:Genetics:
A:Gene: SPDB:SPCC1183.07
A:Map position: 3
C:Superfamily: ribosomal RNA processing protein RRP5

Query Match          35.3%; Score 48; DB 2; Length 1690;
Best Local Similarity 45.5%; Pred. No. 1.2e+02;
Matches 10; Conservative 4; Mismatches 6; Indels 2; Gaps 1;

QY          3  NYSRRLPKGVKHLKDFPILPGE 24
|| | | | | | | | | |
DB          249  NYTGFLSK--KHINDFPFVEGQ 268

RESULT 15
S54011
1-aminocyclopropane-1-carboxylate synthase (EC 4.4.1.14) 2 - moth orchid
C:Species: Doritaenopsis sp. (moth orchid)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 04-Feb-2000
C:Accession: S54011
R:Bui, A.Q.; O'Neill, S.D.
submitted to the EMBL Data Library, January 1993
A:Description: Molecular cloning and characterization of cDNAs encoding 1-aminocyclop
A:Reference number: S54011
A:Accession: S54011
A:Molecule type: mRNA
A:Residues: 1-444 <BUI>
A:Cross-references: EMBL:L07883; NID:gl67980; PIDN:AAB05849.1; PID:gl67981
C:Genetics:
A:Gene: ACS2
C:Superfamily: 1-aminocyclopropane-1-carboxylate synthase
C:Keywords: carbon-sulfur lyase; ethylene biosynthesis; phosphoprotein; pyridoxal p
F;272/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match          34.6%; Score 47; DB 2; Length 444;
Best Local Similarity 62.5%; Pred. No. 42;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY          3  NYSRRLPKGVKHLKDF 18
| | | | | | | | | |
DB          411  NMSRETLEGVKRLKDF 426

Search completed: October 17, 2003, 11:35:57
Job time : 42 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2003, 11:33:06 ; Search time 23 Seconds  
(without alignments)  
51.116 Million cell updates/sec

Title: SEQID2-NAT3  
Perfect score: 136  
Sequence: 1 RPNYSRRLPKGVKHLKDFPILPGEI 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	127	93.4	2351	1	FA8_HUMAN
2	94	69.1	2319	1	FA8_MOUSE
3	75	55.1	2133	1	FA8_PIG
4	54.5	40.1	239	1	UBIG_XANAC
5	54.5	40.1	239	1	UBIG_XANCP
6	52	38.2	599	1	COE4_MOUSE
7	48.5	35.7	405	1	FUT4_HUMAN
8	48.5	35.7	433	1	FUT4_MOUSE
9	48	35.3	495	1	DPO2_RAT
10	48	35.3	600	1	DPO2_MOUSE
11	46	33.8	467	1	GLYD_SCHPO
12	45.5	33.5	246	1	UBIG_XYLFA
13	45.5	33.5	554	1	MHPA_ECOLI
14	45	33.1	438	1	MURA_SYNY3
15	45	33.1	1032	1	ITA4_XENLA
16	45	33.1	1690	1	CA44_HUMAN
17	45	33.1	1691	1	CA64_HUMAN
18	45	33.1	2054	1	YCF2_PINTH
19	44.5	32.7	788	1	SYFB_THEMEA
20	44	32.4	130	1	RR11_MARPO
21	44	32.4	300	1	ERA_MYCTU
22	44	32.4	356	1	CCDA_CAEEL
23	44	32.4	385	1	VATC_METTH
24	44	32.4	1143	1	CALI_HUMAN
25	43.5	32.0	251	1	RPC8_YEAST
26	43.5	32.0	352	1	AROB_SULTO
27	43.5	32.0	649	1	VATI_CHLTR
28	43	31.6	138	1	RR11_PHAAN
29	43	31.6	146	1	YZ37_AQUAE
30	43	31.6	156	1	RS10_LUMRU
31	43	31.6	165	1	RS10_XENLA
32	43	31.6	166	1	RS10_ICTPU
33	43	31.6	227	1	FLPA_PYRHO
					P00451 hmo sapien
					Q06194 mus musculu
					P12263 sus scrofa
					Q8pk00 xanthomonas
					Q8p8h2 xanthomonas
					Q8k4j2 mus musculu
					P220k3 homo sapien
					Q11227 mus musculu
					O89043 rattus norv
					P33611 mus musculu
					O13972 schizosacch
					C9pam5 xylella fas
					P77397 escherichia
					Q55673 synecocyst
					Q91687 xenopus lae
					P53420 homo sapien
					Q14031 homo sapien
					P41653 pinus taunb
					Q9wzs9 thermotoga
					P06364 marchantia
					O05834 mycobacteri
					P35800 caenorhabdi
					Q27038 methanobact
					Q14993 homo sapien
					P17890 saccharomyc
					Q96y96 sulfolobus
					O84307 chlamydia t
					Q8mca0 phaseolus a
					O66425 aquifex aeo
					O77302 lumbricus r
					Q07254 xenopus lae
					Q90yrf4 ictalurus p
					O57811 pyrococcus

34	43	31.6	441	1	CYSP_THEAN	P25781 theileria a
35	43	31.6	529	1	T1M_SALPO	P07989 salmonella
36	43	31.6	529	1	T1M_SALTY	P40813 salmonella
37	43	31.6	579	1	ILVB_SPIPL	P27858 spirulina p
38	43	31.6	689	1	SYM_HALN1	Q9hsa4 halobacteri
39	43	31.6	891	1	SYV_PYRAB	Q9uy55 pyrococcus
40	43	31.6	1082	1	NCO3_RAT	Q9epu2 rattus norv
41	43	31.6	1664	1	RPAL_YEAST	P10964 saccharomyc
42	43	31.6	3856	1	HRX_MOUSE	P55200 mus musculu
43	42.5	31.2	145	1	RS15_XIPMA	P70056 xiphophorus
44	42.5	31.2	188	1	CRI2_HUMAN	P51864 homo sapien
45	42.5	31.2	203	1	RS4_CHLTE	P59129 chlorobium

ALIGNMENTS

RESULT 1					
FA8_HUMAN					
ID	FA8_HUMAN	STANDARD;	PRT;	2351	AA.
AC	P00451;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	Coagulation factor VIII precursor (Procoagulant component)				
DE	(Antihemophilic factor) (AHF).				
GN	F8 OR F8C.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=85081164; PubMed=3935400;				
RA	Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina C.,				
RA	Hartog K., Kuo C.H., Masiarz F.R., Merryweather J.P., Najarian R.,				
RA	Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,				
RA	Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,				
RA	Nordfang O., Ezban M.;				
RT	"Characterization of the polypeptide composition of human factor				
RT	VIII:C and the nucleotide sequence and expression of the human kidney				
RT	CDNA.";				
RL	DNA 4:333-349(1985).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=85061548; PubMed=6438526;				
RA	Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,				
RA	Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,				
RA	Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;				
RA	"Expression of active human factor VIII from recombinant DNA clones.";				
RT	Nature 312:330-337(1984).				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=85061550; PubMed=6438528;				
RA	Toole J.J., Knopf J.L., Wozney J.M., Suizman L.A., Buecker J.L.,				
RA	Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,				
RA	Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,				
RA	Hewick R.M.;				
RT	"Molecular cloning of a cDNA encoding human antihaemophilic factor.";				
RL	Nature 312:342-347(1984).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93265012; PubMed=1303178;				
RA	Gitschier J., Wood W.I.;				
RT	"Sequence of the exon-containing regions of the human factor VIII				
RT	gene.";				
RL	Hum. Mol. Genet. 1:199-200(1992).				
RN	[5]				
RP	SEQUENCE OF 2064-2070 FROM N.A.				
RA	de Water N.S., Williams R., Browett P.J.;				
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.				
RN	[6]				
RP	SULFATION OF TYR-1699.				

RX MEDLINE=91093266; PubMed=1898735;  
RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,  
RA Mertens K., van Mourik J.A.;  
RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is  
RT essential for the interaction of factor VIII with von Willebrand  
RT factor.";  
RL J. Biol. Chem. 266:740-746(1991).  
RN [7]  
RP SULFATION.  
RX MEDLINE=92207952; PubMed=1554716;  
RA Pittman D.D., Wang J.H., Kaufman R.J.;  
RT "Identification and functional importance of tyrosine sulfate  
RT residues within recombinant factor VIII.";  
RL Biochemistry 31:3315-3325(1992).  
RN [8]  
RP---- STRUCTURE BY NMR OF 2322-2343.  
RX MEDLINE=95200924; PubMed=7893714;  
RA Gilbert G.E., Baleja J.D.;  
RT "Membrane-binding peptide from the C2 domain of factor VIII forms an  
RT amphipathic structure as determined by NMR spectroscopy.";  
RL Biochemistry 34:3022-3031(1995).  
RN [9]  
RP REVIEW ON MOLECULAR BASIS OF HEMA.  
RX MEDLINE=91221499; PubMed=1902642;  
RA Gitschier J.;  
RT "The molecular basis of hemophilia A.";  
RL Ann. N.Y. Acad. Sci. 614:89-96(1991).  
RN [10]  
RP REVIEW ON MOLECULAR BASIS OF HEMA.  
RX MEDLINE=89088506; PubMed=2491943;  
RA White G.C. II, Shoemaker C.B.;  
RT "Factor VIII gene and hemophilia A.";  
RL Blood 73:1-12(1989).  
RN [11]  
RP REVIEW ON MOLECULAR BASIS OF HEMA.  
RX MEDLINE=95245332; PubMed=7728145;  
RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;  
RT "Molecular etiology of factor VIII deficiency in hemophilia A.";  
RL Hum. Mutat. 5:1-22(1995).  
RN [12]  
RP VARIANT HEMA GLN-2326.  
RX MEDLINE=86235434; PubMed=3012775;  
RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;  
RT "Identification of a missense mutation in the factor VIII gene of a  
RT mild hemophiliac.";  
RL Science 232:1415-1416(1986).  
RN [13]  
RP VARIANT HEMA PRO-2135.  
RX MEDLINE=88096539; PubMed=3122181;  
RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;  
RT "A novel missense mutation in the factor VIII gene identified by  
RT analysis of amplified hemophilia DNA sequences.";  
RL Nucleic Acids Res. 15:9797-9805(1987).  
RN [14]  
RP VARIANT HEMA GLN-2228.  
RX MEDLINE=88191889; PubMed=2833855;  
RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,  
RA Kazazian H.H.;  
RT "Nonsense and missense mutations in hemophilia A: estimate of the  
RT relative mutation rate at CG dinucleotides.";  
RL Am. J. Hum. Genet. 42:718-725(1988).  
RN [15]  
RP VARIANT HEMA GLY-291.  
RX MEDLINE=88220354; PubMed=2835904;  
RA Youssoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,  
RA Antonarakis S.E.;  
RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution  
RT in exon 7 of the factor VIII gene.";  
RL Am. J. Hum. Genet. 42:867-871(1988).  
RN [16]  
RP VARIANT HEMA CYS-1708.  
RX MEDLINE=89274393; PubMed=2499363;  
RA O'Brien D.P., Tuddenham E.G.;

RT "Purification and characterization of factor VIII 1,689-Cys: a  
RT nonfunctional cofactor occurring in a patient with severe hemophilia  
RT A.";  
RL Blood 73:2117-2122(1989).  
RN [17]  
RP VARIANT HEMA CYS-391.  
RX MEDLINE=90001543; PubMed=2506948;  
RA Shima M., Ware J., Yoshioka A., Fukui H., Fulcher C.A.;  
RT "An arginine to cysteine amino acid substitution at a critical  
RT thrombin cleavage site in a dysfunctional factor VIII molecule.";  
RL Blood 74:1612-1617(1989).  
RN [18]  
RP VARIANT HEMA LEU-189.  
RX MEDLINE=90057680; PubMed=2510835;  
RA Chan V., Chan T.K., Tong T.M., Todd D.;  
RT "A novel missense mutation in exon 4 of the factor VIII:C gene  
RT resulting in moderately severe hemophilia A.";  
RL Blood 74:2688-2691(1989).  
RN [19]  
RP VARIANT HEMA LEU-2326.  
RX MEDLINE=89197216; PubMed=2495245;  
RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;  
RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26  
RT of the factor VIII gene.";  
RL Hum. Genet. 81:335-338(1989).  
RN [20]  
RP VARIANT HEMA HIS-391.  
RX MEDLINE=89264602; PubMed=2498882;  
RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,  
RA Fujimaki M., Hoyer L.W.;  
RT "Direct characterization of factor VIII in plasma: detection of a  
RT mutation altering a thrombin cleavage site  
RT (arginine-372-->histidine).";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).  
RN [21]  
RP VARIANT HEMA CYS-1708.  
RX MEDLINE=90105723; PubMed=2104766;  
RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,  
RA Phillips J.A. III, Janco R.L., Hoyer L.W.;  
RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to  
RT Cys) in the factor VIII gene of two unrelated patients with  
RT cross-reacting material-positive hemophilia A.";  
RL Blood 75:384-389(1990).  
RN [22]  
RP VARIANTS HEMA GLN-2228 AND LEU-2326.  
RX MEDLINE=90123183; PubMed=2105106;  
RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,  
RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;  
RT "Recurrent mutations and three novel rearrangements in the factor  
RT VIII gene of hemophilia A patients of Italian descent.";  
RL Blood 75:662-670(1990).  
RN [23]  
RP VARIANT HEMA CYS-391.  
RX MEDLINE=90329422; PubMed=1973901;  
RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;  
RT "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the  
RT internal heavy chain thrombin cleavage site.";  
RL Br. J. Haematol. 75:73-77(1990).  
RN [24]  
RP VARIANTS HEMA PHE-1699 AND CYS-1708.  
RX MEDLINE=90152691; PubMed=2105906;  
RA Higuchi M., Wong C., Kochhan L., Olek K., Aronis S., Kasper C.K.,  
RA Kazazian H.H., Antonarakis S.E.;  
RT "Characterization of mutations in the factor VIII gene by direct  
RT sequencing of amplified genomic DNA.";  
RL Genomics 6:65-71(1990).  
RN [25]  
RP VARIANTS HEMA CYS-1728 AND ASP-1941.  
RX MEDLINE=90169988; PubMed=2106480;  
RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,  
RA Kazazian H.H.;  
RT "Use of denaturing gradient gel electrophoresis to detect point  
RT mutations in the factor VIII gene.";





RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 705-1573 FROM N.A.  
RX MEDLINE=86287369; PubMed=3016730;  
RA Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C.,  
RA Kaufman R.J.:  
RT "A large region (approximately equal to 95 kDa) of human factor VIII  
RT is dispensable for in vitro procoagulant activity.";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942(1986).  
RN [3]  
RP SEQUENCE OF 392-759 FROM N.A.  
RX MEDLINE=94179260; PubMed=7510693;  
RA Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;  
RT "Elimination of a major inhibitor epitope in factor VIII.";  
RL J. Biol. Chem. 269:8639-8641(1994).  
CC -!- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS  
CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE  
CC ACTIVATED FORM, FACTOR XA.  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.  
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.  
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U49517; AAB06705.1; -.  
DR PIR: A25945; A25945.  
DR PIR: T42763; T42763.  
DR HSSP: P03451; ICFG.  
DR InterPro: IPR001117; Cu-oxidase.  
DR InterPro: IPR000421; FA58\_C.  
DR Pfam: PF00394; Cu-oxidase; 3.  
DR Pfam: PF00754; F5\_F8\_type\_C; 2.  
DR SMART: SM00231; FA58C; 2.  
DR PROSITE; PS01285; FA58C\_1; 2.  
DR PROSITE; PS01286; FA58C\_2; 2.  
DR PROSITE; PS50022; FA58C\_3; 2.  
DR PROSITE; PS00079; MULTICOPPER\_OXIDASE1; 3.  
KW Blood coagulation; Repeat; Plasma; Acute phase; Calcium;  
KW Signal; Glycoprotein; Sulfation.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 2133 COAGULATION FACTOR VIII.  
FT DOMAIN 20 357 F5/8 TYPE A 1.  
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.  
FT DOMAIN 207 357 PLASTOCYANIN-LIKE 2.  
FT DOMAIN 399 730 F5/8 TYPE A 2.  
FT DOMAIN 399 573 PLASTOCYANIN-LIKE 3.  
FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.  
FT DOMAIN 760 1599 B.  
FT DOMAIN 1495 1822 F5/8 TYPE A 3.  
FT DOMAIN 1495 1659 PLASTOCYANIN-LIKE 5.  
FT DOMAIN 1659 1822 PLASTOCYANIN-LIKE 6.  
FT DOMAIN 1822 1970 F5/8 TYPE C 1.  
FT DOMAIN 1975 2127 F5/8 TYPE C 2.  
FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
FT SITE 1449 1450 CLEAVAGE (ACTIVATION) (BY SIMILARITY).  
FT SITE 1490 1491 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).  
FT MOD\_RES 737 737 SULFATION (BY SIMILARITY).  
FT MOD\_RES 738 738 SULFATION (BY SIMILARITY).  
FT MOD\_RES 742 742 SULFATION (BY SIMILARITY).  
FT DISULFID 173 199 PROBABLE.  
FT DISULFID 547 573 PROBABLE.  
FT DISULFID 1633 1659 PROBABLE.  
FT DISULFID 1822 1970 BY SIMILARITY.  
FT DISULFID 1975 2127 BY SIMILARITY.  
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	259	259	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	601	601	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	929	929	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	985	985	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1025	1025	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1111	1111	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1181	1181	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1208	1208	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1245	1245	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1265	1265	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1335	1335	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1408	1408	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1611	1611	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1919	1919	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	713	713	N -> M (IN REF. 2).	
FT	CONFLICT	734	734	I -> T (IN REF. 2).	
FT	CONFLICT	792	792	G -> Q (IN REF. 2).	
FT	CONFLICT	1133	1133	E -> F (IN REF. 2).	
FT	CONFLICT	1191	1191	I -> L (IN REF. 2).	
FT	CONFLICT	1209	1209	R -> F (IN REF. 2).	
FT	CONFLICT	1437	1437	C -> G (IN REF. 2).	
FT	CONFLICT	1456	1456	F -> R (IN REF. 2).	
FT	CONFLICT	1539	1539	F -> R (IN REF. 2).	
FT	CONFLICT	1546	1546	Q -> N (IN REF. 2).	
SQ	SEQUENCE	2133	AA; 239304	MW; 152BBA8997F570DA	CRC64;

Query Match 55.1%; Score 75; DB 1; Length 2133;  
Best Local Similarity 83.3%; Pred. No. 0.0052;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	7	RLPKGVKHLKDFPILPGE	24
DB	509	RLKGVKHLKDMPIPGE	526

RESULT 4  
UBIG\_XANAC  
ID UBIG\_XANAC STANDARD; PRT; 239 AA.  
AC Q8PK00;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 3-demethylubiquinone-9 3-methyltransferase (EC 2.1.1.64) (3,4-  
DE dihydroxy-5-hexaprenylbenzoate methyltransferase) (DHBB  
DE methyltransferase).  
GN UBIG OR XAC2377.  
OS Xanthomonas axonopodis (pv. citri).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=92829;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=306 / ATCC 13902 / XV 101;  
RX MEDLINE=22022145; PubMed=12024217;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,  
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities."  
RL Nature 417:459-463(2002).  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 3-

```
CC demethylubiquinone-9 - S-adenosyl-L-homocysteine + ubiquinone-9.
CC -!- PATHWAY: Ubiquinone biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE UBIG/COQ3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE011874; AAM37229.1; -.
CC HAMAP; MF_00472; -. 1.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR000051; SAM_bind.
KW Ubiquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 239 AA; 26060 MW; F56121516B27DFE6 CRC64;

Query Match 40.1%; Score 54.5; DB 1; Length 239;
Best Local Similarity 57.1%; Pred. No. 0.58;
Matches 12; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

QY 5 SRRLPKGVKHLKDFPILPGEI 25
   :| |||| | ||| | | |
Db 177 ARLLPKGTHHYKDF-IKPAEL 196

RESULT 5
UBIG_XANCP
ID UBIG_XANCP STANDARD; PRT; 239 AA.
AC Q8P8H2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-demethylubiquinone-9 3-methyltransferase (EC 2.1.1.64) (3,4-
DE dihydroxy-5-hexaprenylbenzoate methyltransferase) (DHHB
DE methyltransferase).
DE UBIG OR XCC2269.
GN UBIG
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina J.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moen D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 3-
CC demethylubiquinone-9 - S-adenosyl-L-homocysteine + ubiquinone-9.
CC -!- PATHWAY: Ubiquinone biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE UBIG/COQ3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE012334; AAM41548.1; -.
CC HAMAP; MF_00472; -. 1.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR000051; SAM_bind.
KW Ubiquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 239 AA; 26039 MW; 70BA381BAF14AC71 CRC64;

Query Match 40.1%; Score 54.5; DB 1; Length 239;
Best Local Similarity 57.1%; Pred. No. 0.58;
Matches 12; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

QY 5 SRRLPKGVKHLKDFPILPGEI 25
   :| |||| | ||| | | |
Db 177 ARLLPKGTHHYKDF-IKPSEL 196

RESULT 6
COE4_MOUSE
ID COE4_MOUSE STANDARD; PRT; 599 AA.
AC Q8K4J2; Q8K4J1; Q8K4J3; Q8K4J4; Q8K4J5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor COE4 (Early B-cell factor 4) (EBF-4) (Olf-1/EBF-
DE like 4) (OE-4) (O/E-4).
GN EBF4 OR COE4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND TISSUE
RP SPECIFICITY.
RC STRAIN=C57BL/6;
RX MEDLINE=22136010; PubMed=12139918;
RA Wang S.S., Betz A.G., Reed R.R.;
RT "Cloning of a novel Olf-1/EBF-like gene, O/E-4, by degenerate oligo-
RT based direct selection.";
RL Mol. Cell. Neurosci. 20:404-414(2002).
CC -!- FUNCTION: Seems to weakly activates transcription. Binds an Olf-1
CC consensus site in vitro.
CC -!- SUBUNIT: Forms either a homodimer or a heterodimer with a related
CC family member.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=3; Synonyms=4-23;
CC IsoId=Q8K4J2-1; Sequence=Displayed;
CC Name=1; Synonyms=4-11;
CC IsoId=Q8K4J2-2; Sequence=VSP_001125, VSP_001126;
CC Name=2; Synonyms=4-14;
CC IsoId=Q8K4J2-3; Sequence=VSP_001121, VSP_001122;
CC Name=4; Synonyms=4-132;
CC IsoId=Q8K4J2-4; Sequence=VSP_001123, VSP_001124;
CC Name=5; Synonyms=4S;
CC IsoId=Q8K4J2-5; Sequence=VSP_001119, VSP_001120;
CC -!- TISSUE SPECIFICITY: Expressed in the neuronal and basal cell
CC layers of olfactory epithelium. Absent in the vomeronasal organ.
CC -!- SIMILARITY: BELONGS TO THE COE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```



```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF387630; AAM97580.1; -
DR EMBL; AF387631; AAM97581.1; -
DR EMBL; AF387632; AAM97582.1; -
DR EMBL; AF387633; AAM97583.1; -
DR EMBL; AF387634; AAM97584.1; -
DR MGD; MGI:2385972; Ebf4.
DR GO; GO:0003677; F:DNA binding activity; IDA.
DR GO; GO:0008134; F:transcription factor binding activity; IFI.
DR InterPro; IPR002909; IPT_TIG.
DR InterPro; IPR003523; TF_COE.
DR Pfam; PF01833; TIG; 1.
DR SMART; SM00429; IPT; 1.
DR PROSITE; PS01345; COE; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Activator;
KW Developmental protein; Zinc-finger; Alternative splicing.
FT ZN_FING 152 171 C5-TYPE (POTENTIAL).
FT VARSPLIC 392 426 ELLKRAADVAEALYSAPRAPAPLCPLAPSHPPA -> VW
FT RLCPPPSARGRGSDDPAPAAAPAVPRSLRRSSS (in isoform 5).
FT /FTid=VSP_001119.
FT Missing (in isoform 5).
FT /FTid=VSP_001120.
FT GSYGAPGVTLGVGSPSPFLNGSTATSP -> APRWRLPPP
FT CFFRPppppPASSPSRLST (in isoform 2).
FT /FTid=VSP_001121.
FT Missing (in isoform 2).
FT /FTid=VSP_001122.
FT IMPSSPPLAAASSMSLPAAAPTTSVFSFSPV -> KERLRP
FT CAAPTQFPFIAGLPQSPQSGSRPAF (in isoform 4).
FT /FTid=VSP_001123.
FT Missing (in isoform 4).
FT /FTid=VSP_001124.
FT DQPFED -> AQRTGR (in isoform 1).
FT /FTid=VSP_001125.
FT Missing (in isoform 1).
FT /FTid=VSP_001126.
FT SEQUENCE 599 AA; 64623 MW; 76748B3E04D42260 CRC64;

Query Match 38.2%; Score 52; DB 1; Length 599;
Best Local Similarity 50.0%; Pred. No. 3.9;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 8 LPKGVKHLKDFPILPGFI 25
Db 8 LPRGGLHLKEEPLPSSL 25

RESULT 7
FUT4_HUMAN
ID FUT4_HUMAN STANDARD; PRT; 405 AA.
AC P22083;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-
DE fucosyltransferase) (Fucosyltransferase 4) (FUCT-IV) (ELAM-1 ligand
DE fucosyltransferase).
GN FUT4 OR ELFT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Peripheral blood leukocytes;
RX MEDLINE=91373370; PubMed=1716630;
RA Lowe J.B., Kukowska-Latallo J.F., Nair R.P., Larsen R.D., Marks R.M.,
RA Macher B.A., Kelly R.J., Ernst L.K.;
RT *Molecular cloning of a human fucosyltransferase gene that determines
RT expression of the Lewis x and VIM-2 epitopes but not ELAM-1-dependent
RT cell adhesion.";
```

```
RL J. Biol. Chem. 266:17467-17477(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91084863; PubMed=1702034;
RA Goelz S.E., Hession C., Goff D., Griffiths B., Tizard R., Newman B.,
RA Chi-Rosso G., Lobb R.;
RT *ELFT: a gene that directs the expression of an ELAM-1 ligand.";
RL Cell 63:1349-1356(1990).
RN [3]
RP SEQUENCE OF 1-400 FROM N.A.
RX MEDLINE=92042084; PubMed=1718983;
RA Kumar R., Potvin B., Muller W.A., Stanley P.;
RT "Cloning of a human alpha(1,3)-fucosyltransferase gene that encodes
RT ELFT but does not confer ELAM-1 recognition on Chinese hamster ovary
RT cell transfectants.";
RL J. Biol. Chem. 266:21777-21783(1991).
CC -!- FUNCTION: May catalyze alpha-1,3 glycosidic linkages involved in
CC the expression of Lewis X/SSeA-1 and VIM-2 antigens.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M65030; AAA92977.1; -
DR EMBL; M58596; AAA63172.1; -
DR EMBL; M58597; AAA63173.1; ALT_INIT.
DR EMBL; S65161; AAB20349.1; -
DR PIR; B36340; B36340.
DR Genew; HGNC:4015; FUT4.
DR MIM; 104230; -
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0008417; F:fucosyltransferase activity; TAS.
DR GO; GO:0005975; P:carbohydrate metabolism; TAS.
DR InterPro; IPR001503; Glyco_trans_10.
DR Pfam; PF00852; Glyco_transf_10; 1.
KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;
KW Signal-anchor; Golgi stack.
FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 23 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 48 405 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 87 87 P -> R (IN REF. 2 AND 3).
FT CONFLICT 241 241 E -> D (IN REF. 3).
SQ SEQUENCE 405 AA; 45569 MW; DE72E1FDC390268D CRC64;

Query Match 35.7%; Score 48.5; DB 1; Length 405;
Best Local Similarity 50.0%; Pred. No. 8.6;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 1 RPNYSRRRLPKGV-KHLKDFP 19
Db 319 RANYERFVPRGAFIHVDDFP 338

RESULT 8
FUT4_MOUSE
ID FUT4_MOUSE STANDARD; PRT; 433 AA.
AC Q11127;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-
DE fucosyltransferase) (Fucosyltransferase 4) (FUCT-IV).
```



GN FUT4 OR ELFT.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96027607; Pubmed=7559635;  
RA Gersten K.M., Natsuka S., Trinchera M., Petryniak B., Kelly R.J.,  
RA Hiraiwa N., Jenkins N.A., Gilbert D.J., Copeland N.G., Lowe J.B.;  
RT "Molecular cloning, expression, chromosomal assignment, and tissue-  
RT specific expression of a murine alpha-(1,3)-fucosyltransferase locus  
RT corresponding to the human ELAM-1 ligand fucosyl transferase.";  
RL J. Biol. Chem. 270:25047-25056(1995).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RC STRAIN=129/Sv; TISSUE=Liver;  
RX MEDLINE=97037075; Pubmed=8882722;  
RA Ozawa M., Muramatsu T.;  
RT "Molecular cloning and expression of a mouse alpha-1,3  
RT fucosyltransferase gene that shows homology with the human alpha-1,3  
RT fucosyltransferase IV gene.";  
RL J. Biochem. 119:302-308(1996).  
CC -!- FUNCTION: May catalyze alpha-1,3 glycosidic linkages involved in:  
CC the expression of Lewis X/SSEA-1 and VIM-2 antigens.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
CC form in trans cisternae of Golgi.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long;  
CC IsoId=Q11127-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=Q11127-2; Sequence=VSP\_001778;  
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN STOMACH AND COLON. IT  
CC ALSO EXPRESSED IN THE LUNG, TESTIS, UTERUS, SMALL INTESTINE AND TO  
CC A LESSER EXTENT IN SPLEEN, AND OVARY. PRESENT IN TRACE AMOUNTS IN  
CC BRAIN, THYMUS, HEART, SMOOTH MUSCLE, KIDNEY AND BONE MARROW. NOT  
CC FOUND IN LIVER, SALIVARY GLAND AND PANCREAS.  
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U33457; AAC52269.1; -  
DR EMBL; D63380; BAA09697.1; -  
DR EMBL; D63379; BAA09696.1; -  
DR PIR; A57596; A57596.  
DR MGD; MGI:95594; Fut4.  
DR InterPro; IPR001503; Glyco\_trans\_10.  
DR Pfam; PF00852; Glyco\_transf\_10; 1.  
KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;  
KW Signal-anchor; Golgi stack; Alternative splicing.  
FT DOMAIN 1 52 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 53 74 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT DOMAIN 75 433 LUMENAL, CATALYTIC (POTENTIAL).  
FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 1 33 Missing (in isoform Short).  
FT CONFLICT 252 252 Q -> P (IN REF. 2).  
FT CONFLICT 257 257 R -> Q (IN REF. 2).  
FT CONFLICT 260 260 V -> E (IN REF. 2).  
FT CONFLICT 273 273 R -> Q (IN REF. 2).  
SQ SEQUENCE 433 AA; 49481 MW; 2401822F02B5D021 CRC64;

Query Match 35.7%; Score 48.5; DB 1; Length 433;

Best Local Similarity 50.0%; Pred. No. 9.3;  
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;  
QY 1 RPNYSRRLPKGV-KHLKDFP 19  
| || | :|| | | :|||  
DB 347 RANVERFVPRGAFIHVDDFP 366  
-----  
RESULT 9  
DPO2\_RAT  
ID DPO2\_RAT STANDARD; PRT; 495 AA.  
AC O89043;  
DI 15-JUL-1999 (Rel. 38, Created)  
DI 15-JUL-1999 (Rel. 38, Last sequence update)  
DI 30-MAY-2000 (Rel. 39, Last annotation update)  
DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B;  
DE (DNA polymerase subunit II) (Fragment).  
GN POLA2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RA Popanda O., Flohr C., Thielmann H.W.;  
RT "A mutation in the gene of subunit II of DNA polymerase alpha from  
RT Novikoff cells is concomitant with altered physico-chemical  
RT properties of the enzyme.";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: MAY PLAY AN ESSENTIAL ROLE AT THE EARLY STAGE OF  
CC CHROMOSOMAL DNA REPLICATION BY COUPLING THE POLYMERASE  
CC ALPHA/PRIMASE COMPLEX TO THE CELLULAR REPLICATION MACHINERY (BY  
CC SIMILARITY).  
CC -!- SUBUNIT: DNA POLYMERASE ALPHA-PRIMASE IS A FOUR SUBUNIT ENZYME  
CC (SUBUNITS A, B, C AND D), WHICH IS ASSEMBLED THROUGHOUT THE CELL  
CC CYCLE. THE LARGEST SUBUNIT (SUBUNIT A) HAS DNA POLYMERASE  
CC ACTIVITY, THE TWO SMALLEST SUBUNITS (SUBUNITS C AND D) HAVE DNA  
CC PRIMASE ACTIVITY. SUBUNIT B BINDS TO SUBUNIT A.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M  
CC PHASE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE ALPHA SUBUNIT B FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AJ011606; CAA09721.1; -  
DR Pfam; PF04058; DNA\_pol\_alpha\_B; 1.  
KW DNA replication; Nuclear protein; Phosphorylation.  
FT NON\_TER 1 1  
FT DOMAIN 10 52 PRO/SER/THR-RICH (HYDROPHILIC).  
SQ SEQUENCE 495 AA; 54952 MW; 1816F035737C4A0B CRC64;  
-----  
Query Match 35.3%; Score 48; DB 1; Length 495;  
Best Local Similarity 36.8%; Pred. No. 13;  
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;  
QY 7 RLPKGVKHLKDFPILPGEI 25  
:| | | | :||| :|||:  
DB 178 QIPVDVSELKDYSLFPGQV 196  
-----  
RESULT 10  
DPO2\_MOUSE  
ID DPO2\_MOUSE STANDARD; PRT; 600 AA.  
AC P33611;  
DI 01-FEB-1994 (Rel. 28, Created)









Query Match 33.5%; Score 45.5; DB 1; Length 554;  
Best Local Similarity 41.7%; Pred. No. 35;  
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY 1 RPNYSRRLLPKGVKHLKDFPILPGE 24  
II I II I: :I::II  
Db 220 RPYVSAALPHAVRRF-EFMVMPGE 242

RESULT 14  
MURA\_SYNY3  
ID MURA\_SYNY3 STANDARD; PRT; 438 AA.  
AC Q55573;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)  
DE (Enolpyruvate transferase) (UDP-N-acetylglucosamine enolpyruvyl  
DE transferase) (EPT).  
GN MURA OR MURZ OR SLR0017.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
RA Sugiyama M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
RT region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
CC -!- FUNCTION: CELL WALL FORMATION. ADDS ENOLPYRUVYL TO UDP-N-  
CC ACETYLGLUCOSAMINE (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + UDP-N-acetyl-D-  
CC glucosamine = phosphate + UDP-N-acetyl-3-(1-carboxyvinyl)-D-  
CC glucosamine.  
CC -!- PATHWAY: Peptidoglycan biosynthesis; first step.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- SIMILARITY: Belongs to the EPSP synthase family. Mura subfamily.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: D64000; BAA10199.1; -;  
DR PIR: S76347; S76347.  
DR HSSP: P33038; LDLG.  
DR HAMAP: MF\_00111; -; 1.  
DR InterPro: IPR005750; AcGlu\_Tran\_Mura.  
DR InterPro: IPR001986; EPSP\_synthase.  
DR Pfam: PF00275; EPSP\_synthase; 1.  
DR ProDom: PD001867; EPSP\_synthase; 1.  
DR TIGRFAMs: TIGR01072; mura; 1.  
KW Peptidoglycan synthesis; Cell wall; Cell division; Transferase;  
KW Complete proteome.  
FT ACT\_SITE 129 129 BINDS PEP (BY SIMILARITY).  
SQ SEQUENCE 438 AA; 46604 MW; C310AB287F2F0F55 CRC64;

Query Match 33.1%; Score 45; DB 1; Length 438;  
Best Local Similarity 47.8%; Pred. No. 32;  
Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

QY 5 SRRLLPKGVK--HLKDFPILPGEI 25  
II : II: I II:I I  
Db 228 SRIINGVEKLHSTDFPIPDRI 250

RESULT 15  
ITA4\_XENLA  
ID ITA4\_XENLA STANDARD; PRT; 1032 AA.  
AC Q91687; Q06273;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Integrin alpha-4 precursor (Integrin alpha-IV) (VLA-4).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96379747; PubMed=8787760;  
RA Ramos J.W., Whittaker C.A., Desimone D.W.;  
RT "Integrin-dependent adhesive activity is spatially controlled by  
RT inductive signals at gastrulation.";  
RL Development 122:2873-2883(1996).  
RN [2]  
RP SEQUENCE OF 308-379 FROM N.A.  
RX MEDLINE=94008528; PubMed=8404528;  
RA Whittaker C.A., Desimone D.W.;  
RT "Integrin alpha subunit mRNAs are differentially expressed in early  
RT Xenopus embryos.";  
RL Development 117:1239-1249(1993).  
CC -!- FUNCTION: FIBRONECTIN AND V-CAM ADHESION RECEPTOR.  
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: U54497; AAA98673.1; -;  
DR PIR: I51526; AAA16248.1; -;  
DR HSSP: P11215; IA8X.  
DR InterPro: IPR000413; Integrin\_alpha.  
DR Pfam: PF01839; FG-GAP; 3.  
DR Pfam: PF00357; Integrin\_A; 1.  
DR PRINTS: PR01185; INTEGRINA.  
DR SMART: SM00191; Int\_alpha; 5.  
DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
KW Signal; Repeat.  
FT SIGNAL 1 34 POTENTIAL.  
FT CHAIN 35 1032 INTEGRIN ALPHA-4.  
FT DOMAIN 35 974 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 975 998 POTENTIAL.  
FT DOMAIN 999 1032 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 50 112 FG-GAP 1.  
FT REPEAT ? ? FG-GAP 2.  
FT REPEAT 197 247 FG-GAP 3.  
FT REPEAT 248 301 FG-GAP 4.  
FT REPEAT 303 361 FG-GAP 5.  
FT REPEAT 364 423 FG-GAP 6.  
FT REPEAT 426 474 FG-GAP 7.  
FT CA\_BIND 314 322 POTENTIAL.  
FT CA\_BIND 376 384 POTENTIAL.  
FT CA\_BIND 438 446 POTENTIAL.  
FT SITE 590 591 CLEAVAGE.  
FT SITE 1001 1005 GFFKR MOTIF.  
FT DISULFID 91 101 BY SIMILARITY.  
FT DISULFID 144 165 BY SIMILARITY.  
FT DISULFID 183 198 BY SIMILARITY.

FT	DISULFID	485	494	BY SIMILARITY.
FT	DISULFID	500	556	BY SIMILARITY.
FT	DISULFID	622	627	BY SIMILARITY.
FT	DISULFID	698	712	BY SIMILARITY.
FT	DISULFID	853	889	BY SIMILARITY.
FT	DISULFID	896	901	BY SIMILARITY.
FT	CARBOHYD	81	81	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	98	98	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	229	229	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	479	479	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	496	496	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	517	517	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	537	537	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	626	626	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	660	660	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	746	746	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	857	857	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1032 AA;	115215 MW;	6486797D83AAE69E CRC64;

Query Match 33.1%; Score 45; DB 1; Length 1032;  
Best Local Similarity 52.68; Pred. No. 83;  
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy	7	RLPKGVKHLKDFPILPGEI	25
		: :  :	
Db	678	RLPKGLYFVKVFDLLEKEI	696

Search completed: October 17, 2003, 11:33:39  
Job time : 24 secs





RESULT 2  
 O18806  
 ID O18806 PRELIMINARY; PRI: 2343 AA.  
 AC O18806;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Factor VIII.  
 GN F8.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Cameron C., Notley C., Hoyle S., McGlynn L., Hough C., Kamisue S.,  
 RA Giles A., Lillcrap D.;  
 RT "The canine factor VIII cDNA and 5' flanking sequence.";  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.  
 DR EMBL: AF016234; AAB87412.1; -.  
 DR HSSP: P00451; ICFG.  
 DR InterPro: IPR001117; Cu-oxidase.  
 DR InterPro: IPR000421; FA58\_C.  
 DR Pfam: PF00394; Cu-oxidase; 3.  
 DR Pfam: PF00754; F5\_F8\_type\_C; 2.  
 DR SMART: SM00231; FA58C; 2.  
 DR PROSITE: PS01285; FA58C\_1; 2.  
 DR PROSITE: PS01286; FA58C\_2; 2.  
 DR PROSITE: PS00079; MULTICOPPER\_OXIDASE1; 3.  
 DR SEQUENCE 2343 AA; 265829 MW; A854FAE571C3B399 CRC64;  
 Query Match 74.3%; Score 10.; DB 6; Length 2343;  
 Best Local Similarity 79.2%; Pred. No. 3.3e-06;  
 Matches 19; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PNYSRRLPGVKHLKDFPILPGEI 25  
 : : ||||| ||||| |||||  
 Db 498 PLHTGRZPKGVKHLKDMFPLPGEI 521

RESULT 3  
 Q8EYK5  
 ID Q8EYK5 PRELIMINARY; PRI: 611 AA.  
 AC Q8EYK5;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN LA4208.  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OX NCBI\_TaxID=173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=55601 / Serogroup Icterohaemorrhagiae / Serovar lai;  
 RA Ren S.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE011574; AAN51406.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 611 AA; 69435 MW; 4055F7B88D91594C CRC64;  
 Query Match 39.3%; Score 53.5; DB 16; Length 611;  
 Best Local Similarity 52.2%; Pred. No. 14;  
 Matches 12; Conservative 5; Mismatches 3; Indels 3; Gaps 2;

QY 6 RRLP-KGV-KHLKDFPILPGEI 25  
 : : ||| : : ||||| : :  
 Db 450 KRYPLIKGIFKNVSGFPILPGEV 472

RESULT 4  
 Q8DAC5  
 ID Q8DAC5 PRELIMINARY; PRI: 911 AA.  
 AC Q8DAC5;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE GGDEF family protein.  
 GN VV12281.  
 OS Vibrio vulnificus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of Vibrio vulnificus CMCP6.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE016804; AAO10659.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 911 AA; 103424 MW; 23D2868E85D3AEFD CRC64;  
 Query Match 37.9%; Score 51.5; DB 16; Length 911;  
 Best Local Similarity 43.5%; Pred. No. 44;  
 Matches 10; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

QY 4 YSRRLPGVKHLKD-FPILPGEI 25  
 : ||| : ||| : ||| : |||  
 Db 759 YHARLPDALRHVMDEFDIAPGQL 781

RESULT 5  
 Q96AX6  
 ID Q96AX6 PRELIMINARY; PRI: 166 AA.  
 AC Q96AX6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bladder wait;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC016588; AAH16588.1; -.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 166 AA; 17030 MW; FC6C0918C571C059 CRC64;  
 Query Match 37.5%; Score 51; DB 4; Length 166;  
 Best Local Similarity 36.7%; Pred. No. 7.9;  
 Matches 11; Conservative 5; Mismatches 6; Indels 8; Gaps 1;

QY 2 PNYSRRLPGK-----VKHLKDFPILPG 23  
 : : ||| : : ||| : : ||| : :  
 Db 121 PAYLPRIPPGVRRFAFCIKKLENFPPVPGPG 150

RESULT 6  
 Q8DLH3  
 ID Q8DLH3 PRELIMINARY; PRI: 907 AA.  
 AC Q8DLH3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE T110519 protein.  
 GN T110519.

OS Synechococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=32046;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BP-1;  
RX MEDLINE=22225144; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium  
RT Thermosynechococcus elongatus BP-1.";  
RL DNA Res. 9:123-130(2002).  
DR EMBL; AP005370; BAC08071.1; -.  
KW Complete proteome.  
SQ SEQUENCE 907 AA; 101695 MW; 30DFFCE9B368A775 CRC64;  
  
Query Match 37.5%; Score 51; DB 16; Length 907;  
Best Local Similarity 37.5%; Pred. No. 52;  
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
  
QY 1 RPNYSRRLPKGVKHLKDFPILPGE 24  
||: || :||: ||: ||  
DB 677 RPSLQSRUKNELRHLLPLAPNE 700  
  
RESULT 7  
O65506 SEQUENCE FROM N.A.  
ID O65506 PRELIMINARY; PRT; 1607 AA.  
AC O65506;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative disease resistance protein.  
GN F23E13.30 OR 19K4.270 OR AT4G36140.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Hoheisel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.,  
RA Schueller C.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1448-1607 FROM N.A.  
RA Bevan M., Wedler H., Wambutt R., Hoheisel J., Jesse T., Heijnen L.,  
RA Vos P., Mewes H.W., Mayer K., Schueller C.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1421-1607 FROM N.A.  
RA Wedler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL022141; CAA18120.1; -.  
DR EMBL; AL022373; CAA18508.1; -.  
DR EMBL; AL161588; CAB81523.1; -.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR000767; Disease\_resist.  
DR InterPro; IPR001611; LRR.  
DR InterPro; IPR002182; NB-ARC.  
DR InterPro; IPR000157; TIR\_domain.

DR Pfam; PF00560; LRR; 3.  
DR Pfam; PF00931; NB-ARC; 2.  
DR Pfam; PF01582; TIR; 2.  
DR PRINTS; PR00364; DISEASERSIST.  
DR SMART; SM00382; AAA; 2.  
DR SMART; SM00255; TIR; 2.  
DR PROSITE; PS0104; TIR; 1.  
KW ATP-binding.  
SQ SEQUENCE 1607 AA; 182785 MW; 4F8F572EC72074F3 CRC64;  
  
Query Match 37.5%; Score 51; DB 10; Length 1607;  
Best Local Similarity 66.7%; Pred. No. 97;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 NYSRRLPKGVKHLKD 17  
: || |||||: ||  
DB 573 HYSRLPKGLKFLPD 987  
  
RESULT 8  
Q9S158 PRELIMINARY; PRT; 589 AA.  
AC Q9S158;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE 3-(3-hydroxyphenyl)propionate hydroxylase.  
GN MHPA.  
OS Comamonas testosteroni (Pseudomonas testosteroni).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Comamonadaceae; Comamonas.  
OX NCBI\_TaxID=285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TA441;  
RX MEDLINE=20005595; PubMed=10537203;  
RA Arai H., Yamamoto T., Ohishi T., Shimizu T., Nakata T., Kudo T.;  
RT "Genetic organization and characterization of the 3-(3-  
RT hydroxyphenyl)propionic acid degradation pathway of Comamonas  
RT testosteroni TA441.";  
RL Microbiology 145:2813-2820(1999).  
DR EMBL; AB024335; BAA82878.1; -.  
DR InterPro; IPR001327; FAD\_pyr\_redox.  
DR InterPro; IPR000733; Flav\_monooxygenase.  
DR InterPro; IPR002114; Hpr\_SerP\_site.  
DR InterPro; IPR002938; Mox\_FAD\_binding.  
DR InterPro; IPR000205; NAD\_binding.  
DR InterPro; IPR003042; Rng\_mnxygenase.  
DR InterPro; IPR000594; Thif\_domain.  
DR Pfam; PF01494; FAD\_binding\_3; 1.  
DR Pfam; PF01360; Monooxygenase; 1.  
DR PRINTS; PR00363; FADPNR.  
DR PRINTS; PR00420; RINGMNOXGNASE.  
DR PROSITE; PS00589; PTS\_HPR\_SER; 1.  
SQ SEQUENCE 589 AA; 65297 MW; 7CACF905825CB744 CRC64;  
  
Query Match 37.1%; Score 50.5; DB 2; Length 589;  
Best Local Similarity 41.7%; Pred. No. 38;  
Matches 10; Conservative 5; Mismatches 8; Indels 1; Gaps 1;  
  
QY 1 RPNYSRRLPKGVKHLKDFPILPGE 24  
||: ||: ||: ||: ||: ||  
DB 218 RPYVSAALPHGIRRF-EFMVMPGE 240  
  
RESULT 9  
Q9VE79 PRELIMINARY; PRT; 935 AA.  
ID Q9VE79  
AC Q9VE79;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE CG14309 protein.

GN CG14309.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gaille R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., MCPerson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb C.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Sher H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003721; AAF55548.1;  
 DR FlyBase: FBgn0038611; CG14309.  
 SQ SEQUENCE 935 AA; 106833 MW; 10E46B721CE5DB3F CRC64;

Query Match 37.1%; Score 50.5; DB 5; Length 935;  
 Best Local Similarity 45.8%; Pred. No. 64;  
 Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

QY 2 PNYSRRLPKGVKHLKDFPILPGEI 25  
 |||: |||: |||: |||:  
 Db 575 PTTYT-RLPEGDVHLKHLETVDGEV 597

RESULT 10  
 Q8SAW1 PRELIMINARY; PRT; 388 AA.  
 ID Q8SAW1;  
 AC Q8SAW1;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative wall-associated protein kinase.  
 GN OSJNBA0051J07.1.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,  
 RA Saski C., Henry D., Oates R., Simmons J.;  
 RT "Rice Genomic Sequence";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC098566; AAL77114.1;  
 DR Gramene; Q8SAW1;  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 KW EGF-like domain; Kinase.  
 SQ SEQUENCE 388 AA; 42179 MW; 0F4F47E6F087FF91 CRC64;

Query Match 36.8%; Score 50; DB 10; Length 388;  
 Best Local Similarity 47.6%; Pred. No. 29;  
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 3 NYSRRLPKGVKHLKDFPILPG 23  
 |||: |||: |||: |||:  
 Db 240 NFSKKYPKGVPLVIDFAIRDG 260

RESULT 11  
 Q96A81 PRELIMINARY; PRT; 579 AA.  
 ID Q96A81;  
 AC Q96A81;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein FLJ31952.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ninomiya K., Waqatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa I., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isofai T.;  
 RT "NEDO human cDNA sequencing project";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK056514; BAB71201.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 579 AA; 65949 MW; 195AEFB2F4235C84 CRC64;

Query Match 36.8%; Score 50; DB 4; Length 579;  
 Best Local Similarity 40.0%; Pred. No. 45;  
 Matches 12; Conservative 5; Mismatches 5; Indels 8; Gaps 2;

QY 1 RPNYSRRLPKGVKHLKD-----FPILPGEI 25  
 |||: |||: |||: |||: |||: |||:  
 Db 61 RPVYSK---KGLEHKADLQHLFPVPPGHL 87

RESULT 12  
 Q8S5J1 PRELIMINARY; PRT; 706 AA.  
 ID Q8S5J1;  
 AC Q8S5J1;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative wall-associated protein kinase.  
 GN OSJNBA0011A24.33.





RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
[2]  
RN SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[4]  
RN SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
RN SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
[6]  
RN SEQUENCE FROM N.A.  
RA Boube M., Faucher C., Joulia L., Cribbs D.L., Bourbon H.M.;  
RT "The Drosophila Mediator proteins DTRAP240 and DTRAP80 are  
RT transcriptional cofactors of the proboscipedia and Sex combs Reduced  
RT homeotic proteins.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
[7]  
RN SEQUENCE FROM N.A.  
RA Nairz K., Hafen E.;  
RT "Isolation of flytrap (pap), the Drosophila TRAP240 homologue.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
[8]  
RN SEQUENCE FROM N.A.  
RA Boube M., Faucher C., Joulia L., Cribbs D.L., Bourbon H.M.;  
RT "The Drosophila Mediator proteins DTRAP240 and DTRAP80 are  
RT transcriptional cofactors of the proboscipedia and Sex combs reduced  
RT homeotic proteins.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
[9]  
RN SEQUENCE FROM N.A.  
RA Treisman J.E.;  
RT "Drosophila homologs of the transcriptional coactivation complex  
RT subunits TRAP240 and TRAP230 are required for identical processes in  
RT eye-antennal disc development.";  
RL Development 0:0-0(2001).  
DR EMBL; AE003593; AAF51667.2; -.

DR EMBL; AF227215; AAF43172.1; -.  
DR EMBL; AF226855; AAF36691.1; -.  
DR EMBL; AF227214; AAF43021.1; -.  
DR EMBL; AF324425; AAC48327.1; -.  
DR FlyBase; FBgn0024200; pap.  
SQ SEQUENCE 2618 AA; 280021 MW; 735A8A502076844E CRC64;  
  
Query Match 36.8%; Score 50; DB 5; Length 2618;  
Best Local Similarity 40.9%; Pred. NO. 2.4e+02;  
Matches 9; Conservative 6; Mismatches 5; Indels 2; Gaps 1;  
  
QY 1 RPNYSRRRLPKGVKHLKDFPILP 22  
Db 836 RPRYAKNLYEQNHVK--PVMP 855  
  
RESULT 15  
O58994 PRELIMINARY; PRT; 412 AA.  
AC O58994;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Hypothetical protein PH1259.  
GN PH1259.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.;  
RI "Complete sequence and gene organization of the genome of a hyper-  
RI thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
DR EMBL; AP000005; BAA30361.1; -.  
DR InterPro; IPR000051; SAM\_bind.  
DR InterPro; IPR001566; Tima.  
DR PROSITE; PS01231; TRMA\_2; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 412 AA; 47094 MW; FFB2CDD11EF5B331 CRC64;  
  
Query Match 36.0%; Score 49; DB 17; Length 412;  
Best Local Similarity 58.8%; Pred. NO. 44;  
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 2 PNYSRRLPKGVKHLKDF 18  
Db 140 PIFSERTPKYLKALKDF 156  
  
Search completed: October 17, 2003, 11:37:41  
Job time : 99 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run On: October 17, 2003, 11:33:45 ; Search time 30 seconds  
(without alignments);  
35.259 Million cell updates/sec

Title: SEQID2-NAT3  
Perfect score: 136  
Sequence: 1 RPNYSRRLPKGVKHLKDFPILPGEI 25

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310856 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score		Query Match	Length	DB ID	Description
		%				
1	127	93.4	1438	4	US-09-209-916-1	Sequence 1, Appli
2	127	93.4	1471	1	US-08-683-839B-3	Sequence 3, Appli
3	127	93.4	1661	2	US-08-882-083-2	Sequence 2, Appli
4	127	93.4	1661	2	US-08-558-107-2	Sequence 2, Appli
5	127	93.4	1661	3	US-09-243-539-2	Sequence 2, Appli
6	127	93.4	2332	1	US-07-864-004B-4	Sequence 4, Appli
7	127	93.4	2332	1	US-08-251-937A-4	Sequence 4, Appli
8	127	93.4	2332	1	US-08-212-133A-2	Sequence 2, Appli
9	127	93.4	2332	1	US-08-276-594A-2	Sequence 2, Appli
10	127	93.4	2332	1	US-08-474-503-2	Sequence 2, Appli
11	127	93.4	2332	2	US-08-670-707A-2	Sequence 2, Appli
12	127	93.4	2332	3	US-09-037-601-2	Sequence 2, Appli
13	127	93.4	2332	3	US-09-324-867-3	Sequence 3, Appli
14	127	93.4	2332	4	US-09-315-179-2	Sequence 2, Appli
15	127	93.4	2332	4	US-09-523-656-2	Sequence 2, Appli
16	127	93.4	2332	5	PCT-US93-03275-4	Sequence 4, Appli
17	127	93.4	2332	5	PCT-US94-13200-2	Sequence 2, Appli
18	127	93.4	2351	1	US-08-121-202-2	Sequence 2, Appli
19	127	93.4	2351	1	US-08-366-851A-2	Sequence 2, Appli
20	127	93.4	2351	6	5171844-2	Patent No. 517:844
21	127	93.4	2351	6	5422260-1	Patent No. 5422260
22	101	74.3	2343	3	US-09-324-867-2	Sequence 2, Appli
23	94	69.1	2304	3	US-09-324-867-4	Sequence 4, Appli
24	94	69.1	2319	1	US-08-212-133A-8	Sequence 8, Appli
25	94	69.1	2319	1	US-08-474-503-6	Sequence 6, Appli
26	94	69.1	2319	2	US-08-670-707A-6	Sequence 6, Appli
27	94	69.1	2319	3	US-09-037-601-6	Sequence 6, Appli

28	94	69.1	2319	4	US-09-315-179-6	Sequence 6, Appli
29	94	69.1	2319	4	US-09-523-656-28	Sequence 28, Appli
30	94	69.1	2319	5	PCT-US94-13200-6	Sequence 6, Appli
31	75	55.1	367	1	US-07-864-004B-2	Sequence 2, Appli
32	75	55.1	367	1	US-08-251-937A-2	Sequence 2, Appli
33	75	55.1	367	5	PCT-US93-03275-2	Sequence 2, Appli
34	75	55.1	368	1	US-08-212-133A-6	Sequence 6, Appli
35	75	55.1	368	1	US-08-474-503-4	Sequence 4, Appli
36	75	55.1	368	2	US-08-670-707A-4	Sequence 4, Appli
37	75	55.1	368	3	US-09-037-601-4	Sequence 4, Appli
38	75	55.1	368	4	US-09-315-179-4	Sequence 4, Appli
39	75	55.1	368	4	US-09-523-656-4	Sequence 4, Appli
40	75	55.1	368	5	PCT-US94-13200-4	Sequence 4, Appli
41	75	55.1	541	1	US-08-121-202-4	Sequence 4, Appli
42	75	55.1	1443	2	US-08-670-707A-39	Sequence 39, Appli
43	75	55.1	1443	3	US-09-037-601-39	Sequence 39, Appli
44	75	55.1	1443	4	US-09-315-179-39	Sequence 39, Appli
45	75	55.1	1467	4	US-09-523-656-38	Sequence 38, Appli

ALIGNMENTS

RESULT 1  
US-09-209-916-1  
; Sequence 1, Application US/09209916  
; Patent No. 6358703  
; GENERAL INFORMATION:  
; APPLICANT: Cho, Myung-Sam  
; APPLICANT: Chan, Sham-Yuen  
; APPLICANT: Kelsey, William  
; APPLICANT: Yee, Helena  
; TITLE OF INVENTION: Expression System for Factor VIII  
; FILE REFERENCE: MSB-7255  
; CURRENT APPLICATION NUMBER: US/09/209,916  
; CURRENT FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1438  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Derived from  
; OTHER INFORMATION: human factor VIII sequence  
US-09-209-916-1

Query Match 93.4%; Score 127; DB 4; Length 1438;  
Best Local Similarity 96.0%; Pred. No. 4.3e-11;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPKGVKHLKDFPILPGEI 25  
|||  
Db 484 RPNYSRRLPKGVKHLKDFPILPGEI 508

RESULT 2  
US-08-683-839B-3  
; Sequence 3, Application US/08683839B  
; Patent No. 5744326  
; GENERAL INFORMATION:  
; APPLICANT: Ill, Charles . R. et al.  
; TITLE OF INVENTION: Use of Viral Cis-Acting Post-Transcriptional  
; TITLE OF INVENTION: Regulatory Sequences To Increase Expression of  
; TITLE OF INVENTION: Intronsless Genes Containing Near-Consensus Splice Sites  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, suite 510  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/683,839B  
FILING DATE: 11-MARCH-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Remillard, Jane E.  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: TTI-138  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1471 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-683-839B-3

Query Match 93.4%; Score 127; DB 1; Length 1471;  
Best Local Similarity 96.0%; Pred. No. 4.4e-11;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25  
|| |||||  
Db 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

RESULT 3  
US-08-882-083-2  
Sequence 2, Application US/08882083  
Patent No. 5869292  
GENERAL INFORMATION:  
APPLICANT: VOORBERG, Johannes J.  
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/882,083  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/558,107  
FILING DATE: 13-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ISACSON, John P.  
REGISTRATION NUMBER: 33,715  
REFERENCE/DOCKET NUMBER: 30472/212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1661 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-882-083-2  
Query Match 93.4%; Score 127; DB 2; Length 1661;  
Best Local Similarity 96.0%; Pred. No. 5.1e-11;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25  
|| |||||  
Db 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

RESULT 4  
US-08-558-107-2  
Sequence 2, Application US/08558107  
Patent No. 5910481  
GENERAL INFORMATION:  
APPLICANT: VOORBERG, Johannes J.  
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/558,107  
FILING DATE: 13-NOV-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ISACSON, John P.  
REGISTRATION NUMBER: 33,715  
REFERENCE/DOCKET NUMBER: 30472/212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1661 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-558-107-2

Query Match 93.4%; Score 127; DB 2; Length 1661;  
Best Local Similarity 96.0%; Pred. No. 5.1e-11;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25  
|| |||||  
Db 503 RPLYSRRLPKGVKHLKDFPILPGEI 527

RESULT 5  
US-09-243-539-2  
Sequence 2, Application US/09243539  
Patent No. 6130203  
GENERAL INFORMATION:  
APPLICANT: VOORBERG, Johannes J.  
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:



```
; TISSUE TYPE: Liver cDNA sequence
US-08-251-937A-4

Query Match          93.4%; Score 127; DB 1; Length 2332;
Best Local Similarity 96.0%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
   || ||||| ||||| ||||| ||||| |||||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 8
US-08-212-133A-2
; Sequence 2, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marschall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,133A
; FILING DATE: March 11, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: FMU/76677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
US-08-212-133A-2

Query Match          93.4%; Score 127; DB 1; Length 2332;
Best Local Similarity 96.0%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
   || ||||| ||||| ||||| ||||| |||||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 9
US-08-276-594A-2
```

```
; Sequence 2, Application US/08276594A
; Patent No. 5693499
; GENERAL INFORMATION:
; APPLICANT: YONEMURA, Hiroshi
; APPLICANT: TAJIMA, Yoshitaka
; APPLICANT: SUGAWARA, Keishin
; APPLICANT: MASUDA, Kenichi
; TITLE OF INVENTION: PROCESS FOR PREPARING HUMAN COAGULATION
; TITLE OF INVENTION: FACTOR VIII PROTEIN COMPLEX
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,594A
; FILING DATE: 18-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/950,191
; FILING DATE: 24-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 243262/1991
; FILING DATE: 24-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/195/AOPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-276-594A-2

Query Match          93.4%; Score 127; DB 1; Length 2332;
Best Local Similarity 96.0%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
   || ||||| ||||| ||||| ||||| |||||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 10
US-08-474-503-2
; Sequence 2, Application US/08474503
; Patent No. 5744446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
```



; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
; US-08-474-503-2

Query Match 93.4%; Score 127; DB 1; Length 2332;
Best Local Similarity 96.0%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 11
US-08-670-707A-2
; Sequence 2, Application US/08670707A
; Patent No. 5859204
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,707A
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:

; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; US-08-670-707A-2

Query Match 93.4%; Score 127; DB 2; Length 2332;
Best Local Similarity 96.0%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRLPKGVKHLKDFPILPGEI 25
|| |||||
Db 484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 12
US-09-037-601-2
; Sequence 2, Application US/09037601
; Patent No. 6180371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037.601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
US-09-037-601-2

Query Match          93.4%; Score 127; DB 3; Length 2332;
Best Local Similarity 96.0%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db      484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 13
US-09-324-867-3
; Sequence 3, Application US/09324867A
; Patent No. 6251632
; GENERAL INFORMATION:
; APPLICANT: Lilligrap, David
; APPLICANT: Cameron, Cherie
; APPLICANT: No. 6251632ley, Colleen
; APPLICANT: Horrocks, L. Suzanne Hoyle
; APPLICANT: Hough, Christine
; TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
; FILE REFERENCE: 1669.0010002/JAG/BJD
; CURRENT APPLICATION NUMBER: US/09/324.867A
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: 09/035,141
; EARLIER FILING DATE: 1998-03-059
; EARLIER APPLICATION NUMBER: 60/039,953
; EARLIER FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-324-867-3

Query Match          93.4%; Score 127; DB 3; Length 2332;
Best Local Similarity 96.0%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db      485 RPLYSRRLPKGVKHLKDFPILPGEI 509

RESULT 14
US-09-315-179-2
; Sequence 2, Application US/09315179
; Patent No. 6376463
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95H
; CURRENT APPLICATION NUMBER: US/09/315,179
; CURRENT FILING DATE: 1999-05-20
; EARLIER APPLICATION NUMBER: U.S. 09/037,601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: U.S. 08/670,707
; EARLIER FILING DATE: 1996-06-26
; EARLIER APPLICATION NUMBER: PCT/US97/11155
; EARLIER FILING DATE: 1997-06-26
; EARLIER APPLICATION NUMBER: PCT/US94/13200

```

```

; EARLIER FILING DATE: 1994-11-15
; EARLIER APPLICATION NUMBER: U.S. 08/212,133
; EARLIER FILING DATE: 1994-03-11
; EARLIER APPLICATION NUMBER: U.S. 07/864,004
; EARLIER FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-179-2

Query Match          93.4%; Score 127; DB 4; Length 2332;
Best Local Similarity 96.0%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db      484 RPLYSRRLPKGVKHLKDFPILPGEI 508

RESULT 15
US-09-523-656-2
; Sequence 2, Application US/09523656
; Patent No. 6458563
; GENERAL INFORMATION:
; APPLICANT: Lollar S., John
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-95I
; CURRENT APPLICATION NUMBER: US/09/523,656
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 09/037,601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: 08/670,707
; EARLIER FILING DATE: 1996-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-523-656-2

Query Match          93.4%; Score 127; DB 4; Length 2332;
Best Local Similarity 96.0%; Pred. No. 7.4e-11;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 RPNYSRRLPKGVKHLKDFPILPGEI 25
Db      484 RPLYSRRLPKGVKHLKDFPILPGEI 508

Search completed: October 17, 2003, 11:45:06
Job time : 31 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OMprotein - protein search, using sw model  
Run on: October 17, 2003, 11:36:06 ; Search time 59 Seconds  
(without alignments)  
59.390 Million cell updates/sec

Title: SEQID2-NAT3  
Perfect score: 136  
Sequence: 1 RPNYSRRRLPGVKHLKDFPILPGEI 25  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 609560 seqs, 163917102 residues  
Total number of hits satisfying chosen parameters: 609560  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA: \*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	127	93.4		1438	14	US-10-006-091-1	Sequence 1, Appli
2	127	93.4		1438	14	US-10-047-257-1	Sequence 1, Appli
3	127	93.4		1438	15	US-10-225-900-1	Sequence 1, Appli
4	127	93.4		1471	14	US-10-095-718-2	Sequence 2, Appli
5	127	93.4		2332	10	US-09-957-641-2	Sequence 2, Appli
6	127	93.4		2332	12	US-10-131-510A-2	Sequence 2, Appli
7	127	93.4		2332	15	US-10-187-319-2	Sequence 2, Appli
8	127	93.4		2351	15	US-10-132-829-4	Sequence 4, Appli
9	127	93.4		2351	15	US-10-172-712-27	Sequence 27, Appli
10	101	74.3		1431	14	US-10-095-718-4	Sequence 4, Appli
11	94	69.1		2319	12	US-10-131-510A-6	Sequence 6, Appli
12	94	69.1		2319	15	US-10-187-319-6	GENERAL INFORMA
13	75	55.1		368	12	US-10-131-510A-4	Sequence 4, Appli
14	75	55.1		368	15	US-10-187-319-4	Sequence 4, Appli
15	75	55.1		1443	12	US-10-131-510A-39	Sequence 39, Appli

16	75	55.1	1443	15	US-10-187-319-39	Sequence 39, Appli
17	75	55.1	2133	12	US-10-131-510A-37	Sequence 37, Appli
18	75	55.1	2133	15	US-10-187-319-37	Sequence 37, Appli
19	50	36.8	897	10	US-09-949-192-49	Sequence 49, Appli
20	48.5	35.7	405	10	US-09-863-475A-8	Sequence 8, Appli
21	48.5	35.7	405	12	US-10-234-041-11	Sequence 11, Appli
22	48.5	35.7	432	9	US-09-733-524-7	Sequence 7, Appli
23	48.5	35.7	433	12	US-10-189-977-11	Sequence 11, Appli
24	48.5	35.7	433	12	US-10-392-098-11	Sequence 11, Appli
25	48.5	35.7	433	14	US-10-120-319-11	Sequence 11, Appli
26	48.5	35.7	530	12	US-10-234-041-13	Sequence 13, Appli
27	48	35.3	499	15	US-10-156-761-11906	Sequence 11906, A
28	47	34.6	345	10	US-09-738-626-6584	Sequence 6584, Ap
29	47	34.6	826	11	US-09-988-626-2	Sequence 2, Appli
30	47	34.6	826	11	US-09-988-687-2	Sequence 2, Appli
31	47	34.6	826	11	US-09-988-686-2	Sequence 2, Appli
32	45	33.1	444	15	US-10-157-223-8	Sequence 8, Appli
33	44.5	32.7	331	12	US-10-278-946-18	Sequence 18, Appli
34	44.5	32.7	376	10	US-09-834-722-2	Sequence 2, Appli
35	44.5	32.7	441	10	US-09-738-626-4005	Sequence 4005, Ap
36	44	32.4	282	9	US-09-893-637-2	Sequence 2, Appli
37	44	32.4	282	9	US-09-970-711-21	Sequence 21, Appli
38	44	32.4	388	15	US-10-156-761-14067	Sequence 14067, A
39	44	32.4	482	9	US-09-796-858-40	Sequence 40, Appli
40	44	32.4	491	10	US-09-997-664-3	Sequence 3, Appli
41	44	32.4	826	11	US-09-988-626-226	Sequence 226, App
42	44	32.4	826	11	US-09-988-687-226	Sequence 226, App
43	44	32.4	826	11	US-09-988-686-226	Sequence 226, App
44	44	32.4	913	12	US-09-976-782-94	Sequence 94, Appli
45	44	32.4	1142	12	US-09-976-782-92	Sequence 92, Appli

ALIGNMENTS

RESULT 1  
US-10-006-091-1  
; Sequence 1, Application US/10006091  
; Publication No. US20020102730A1  
; GENERAL INFORMATION:  
; APPLICANT: Cho, Myung-Sam  
; APPLICANT: Chan, Sham-Yuen  
; APPLICANT: Kelsey, William  
; APPLICANT: Yee, Helena  
; TITLE OF INVENTION: Expression System for Factor VIII  
; FILE REFERENCE: MSB-7255.1  
; CURRENT APPLICATION NUMBER: US/10/006,091  
; CURRENT FILING DATE: 2001-12-06  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1438  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Derived from  
; OTHER INFORMATION: human factor VIII sequence  
US-10-006-091-1

Query Match 93.4%; Score 127; DB 14; Length 1438;  
Best Local Similarity 96.0%; Pred. No. 1.3e-10;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPNYSRRRLPGVKHLKDFPILPGEI 25  
|||  
Db 484 RPLYSRRRLPGVKHLKDFPILPGEI 508

RESULT 2  
US-i0-047-257-1  
; Sequence 1, Application US/10047257  
; Publication No. US20020115152A1  
; GENERAL INFORMATION:



```

; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.2
; CURRENT APPLICATION NUMBER: US/10/047,257
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-047-257-1
```

```

Query Match          93.4%;   Score 127;   DB 14;   Length 1438;
Best Local Similarity 96.0%;   Pred. No. 1.3e-10;
Matches 24;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;
```

```
Qy      1 RPNYSRRRLPGVKVHKLKDFPILPGEI 25
      || ||||| ||||| ||||| ||||| |||||
Db      484 RPLYSRRRLPGVKVHKLKDFPILPGEI 508
```

```

RESULT 3
US-10-225-900-1
; Sequence 1, Application US/10225900
; Publication No. US20030077752A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255
; CURRENT APPLICATION NUMBER: US/10/225,900
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US/09/209,916
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-225-900-1
```

```

Query Match          93.4%;   Score 127;   DB 15;   Length 1438;
Best Local Similarity 96.0%;   Pred. No. 1.3e-10;
Matches 24;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;
```

```
Qy      1 RPNYSRRRLPGVKVHKLKDFPILPGEI 25
      || ||||| ||||| ||||| ||||| |||||
Db      484 RPLYSRRRLPGVKVHKLKDFPILPGEI 508
```

```

RESULT 4
US-10-095-718-2
; Sequence 2, Application US/10095718
; Publication No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
```

```

; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: PRT
; ORGANISM: Homo sapiens B-domain deleted factor VIII
; FEATURE:
; OTHER INFORMATION: Homo sapiens BDD FVIII
US-10-095-718-2
```

```

Query Match          93.4%;   Score 127;   DB 14;   Length 1471;
Best Local Similarity 96.0%;   Pred. No. 1.4e-10;
Matches 24;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;
```

```
Qy      1 RPNYSRRRLPGVKVHKLKDFPILPGEI 25
      || ||||| ||||| ||||| ||||| |||||
Db      503 RPLYSRRRLPGVKVHKLKDFPILPGEI 527
```

```

RESULT 5
US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; CURRENT FILING DATE: 2001-09-16
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-641-2
```

```

Query Match          93.4%;   Score 127;   DB 10;   Length 2332;
Best Local Similarity 96.0%;   Pred. No. 2.3e-10;
Matches 24;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;
```

```
Qy      1 RPNYSRRRLPGVKVHKLKDFPILPGEI 25
      || ||||| ||||| ||||| ||||| |||||
Db      484 RPLYSRRRLPGVKVHKLKDFPILPGEI 508
```

```

RESULT 6
US-10-131-510A-2
; Sequence 2, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
```

; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-510A-2

Query Match 93.4%; Score 127; DB 12; Length 2332;
Best Local Similarity 96.0%; Pred. No. 2.3e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPGVKVHLKDFPILPGEI 25
Db 484 RPLYSRRLPGVKVHLKDFPILPGEI 508

RESULT 7
US-10-187-319-2
; Sequence 2, Application US/10187319
; Publication No. US20030068785A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/187,319
; FILING DATE: 27-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/523,656
; FILING DATE: 2000-03-10
; APPLICATION NUMBER: US 09/037,601
; FILING DATE: 1998-03-10
; APPLICATION NUMBER: WC PCT/US97/11155
; FILING DATE: 1997-06-26
; APPLICATION NUMBER: US 08/670,707
; FILING DATE: 1996-06-26
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: <Unknown>
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-187-319-2

Query Match 93.4%; Score 127; DB 15; Length 2332;
Best Local Similarity 96.0%; Pred. No. 2.3e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPGVKVHLKDFPILPGEI 25
Db 484 RPLYSRRLPGVKVHLKDFPILPGEI 508

RESULT 8
US-10-132-829-4
; Sequence 4, Application US/10132829
; Publication No. US20030044982A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor
; TITLE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-PA1170
; CURRENT APPLICATION NUMBER: US/10/132,829
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/286,314
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-132-829-4

Query Match 93.4%; Score 127; DB 15; Length 2351;
Best Local Similarity 96.0%; Pred. No. 2.3e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPNYSRRLPGVKVHLKDFPILPGEI 25
Db 503 RPLYSRRLPGVKVHLKDFPILPGEI 527

RESULT 9
US-10-172-712-27
; Sequence 27, Application US/10172712
; Publication No. US20030125232A1
; GENERAL INFORMATION:
; APPLICANT: GRIFFIN, JOHN H.
; APPLICANT: GALE, ANDREW J.
; APPLICANT: GETZOFF, ELIZABETH D.
; APPLICANT: PELLEQUER, JEAN-LUC
; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS
; FILE REFERENCE: 4198-4001US1
; CURRENT APPLICATION NUMBER: US/10/172,712
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: 60/298,578
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-172-712-27





Sequence 4, Application US/10131510A  
Publication No. US20030166536A1  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S  
TITLE OF INVENTION: Modified Factor VIII  
FILE REFERENCE: 75-95J  
CURRENT APPLICATION NUMBER: US/10/131,510A  
CURRENT FILING DATE: 2002-07-10  
PRIOR APPLICATION NUMBER: U.S. 09/315,179  
PRIOR FILING DATE: 1999-05-20  
PRIOR APPLICATION NUMBER: U.S. 09/037,601  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: U.S. 08/670,707  
PRIOR FILING DATE: 1996-06-26  
PRIOR APPLICATION NUMBER: PCT/US97/11155  
PRIOR FILING DATE: 1997-06-26  
PRIOR APPLICATION NUMBER: PCT/US94/13200  
PRIOR FILING DATE: 1994-11-15  
PRIOR APPLICATION NUMBER: U.S. 08/212,133  
PRIOR FILING DATE: 1994-03-11  
PRIOR APPLICATION NUMBER: U.S. 07/864,004  
PRIOR FILING DATE: 1992-04-07  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 368  
TYPE: PRT  
ORGANISM: Sus scrofa  
US-10-131-510A-4

Query Match 55.1%; Score 75; DB 12; Length 368;  
Best Local Similarity 83.3%; Pred. No. 0.0027;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 RLPKGVKHLKDFPILPGE 24  
|| || ||||| |||||  
Db 118 RLKGVKHLKDMPILPGE 135

RESULT 14  
US-10-187-319-4  
Sequence 4, Application US/10187319  
Publication No. US20030068785A1  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S.  
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/187,319  
FILING DATE: 27-Aug-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/523,656  
FILING DATE: 2000-03-10  
APPLICATION NUMBER: US 09/037,601  
FILING DATE: 1998-03-10  
APPLICATION NUMBER: WO PCT/US97/11155  
FILING DATE: 1997-06-26  
APPLICATION NUMBER: US 08/670,707  
FILING DATE: 1996-06-26  
ATTORNEY/AGENT INFORMATION:

NAME: Greenlee, Lorraine L.  
REGISTRATION NUMBER: 27,894  
REFERENCE/DOCKET NUMBER: 75-95K  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 368 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Porcine  
TISSUE TYPE: spleen  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..368  
OTHER INFORMATION: /note= "Predicted amino acid  
sequence of porcine factor VIII A2 domain, defined as residues  
homologous to human factor VIII, amino acids 373-740. Residues  
1-4 are from known porcine amino acid sequence."  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-187-319-4

Query Match 55.1%; Score 75; DB 15; Length 368;  
Best Local Similarity 83.3%; Pred. No. 0.0027;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 RLPKGVKHLKDFPILPGE 24  
|| || ||||| |||||  
Db 118 RLKGVKHLKDMPILPGE 135

RESULT 15  
US-10-131-510A-39  
Sequence 39, Application US/10131510A  
Publication No. US20030166536A1  
GENERAL INFORMATION:  
APPLICANT: Lollar, John S  
TITLE OF INVENTION: Modified Factor VIII  
FILE REFERENCE: 75-95J  
CURRENT APPLICATION NUMBER: US/10/131,510A  
CURRENT FILING DATE: 2002-07-10  
PRIOR APPLICATION NUMBER: U.S. 09/315,179  
PRIOR FILING DATE: 1999-05-20  
PRIOR APPLICATION NUMBER: U.S. 09/037,601  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: U.S. 08/670,707  
PRIOR FILING DATE: 1996-06-26  
PRIOR APPLICATION NUMBER: PCT/US97/11155  
PRIOR FILING DATE: 1997-06-26  
PRIOR APPLICATION NUMBER: PCT/US94/13200  
PRIOR FILING DATE: 1994-11-15  
PRIOR APPLICATION NUMBER: U.S. 08/212,133  
PRIOR FILING DATE: 1994-03-11  
PRIOR APPLICATION NUMBER: U.S. 07/864,004  
PRIOR FILING DATE: 1992-04-07  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 39  
LENGTH: 1443  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: factor VIII lacking  
OTHER INFORMATION: the B domain  
US-10-131-510A-39

Query Match 55.18; Score 75; DB 12; Length 1443;  
Best Local Similarity 83.38; Pred. NO. 0.012;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 7 RLPKGVKHLKDFPILPGE 24  
||| |||||  
Db 509 RLLKGWKHLKOMPILPGE 526

Search completed: October 17, 2003, 11:46:54  
Job time : 70 secs